

(19) World Intellectual Property
Organization
International Bureau



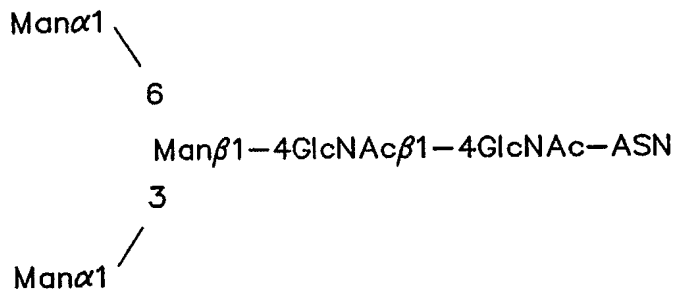
(43) International Publication Date
24 November 2005 (24.11.2005)

PCT

(10) International Publication Number
WO 2005/111627 A2

- (51) International Patent Classification⁷: **G01N 33/68**
- (74) Agent: **VATLAND, Janice, A.**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).
- (21) International Application Number:
PCT/US2005/013107
- (81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (22) International Filing Date: 15 April 2005 (15.04.2005)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/562,874 15 April 2004 (15.04.2004) US
- (71) Applicant (*for all designated States except US*): **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **KEISER, Nishla** [US/US]; 46 Lopez Street, Cambridge, MA 02141 (US). **BOSQUES, Carlos** [US/US]; 71 Fulkerson Street, Cambridge, MA 02141 (US). **SASISEKHARAN, Ram** [US/US]; 4 Duval Way, Bedford, MA 01730 (US). **GANDHE, Pankaj** [IN/US]; 91 Windingwood Drive, Apt. 6B, Sayreville, NJ 08872 (US). **RAGURAM, S.** [US/US]; 34 Vliet Drive, Hillsborough, NJ 08844 (US). **SRINIVASAN, Aravind** [IN/US]; 99 Florence Street, #416, Malden, MA 02148 (US).
- (84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— *without international search report and to be republished upon receipt of that report*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: METHODS AND PRODUCTS RELATED TO THE IMPROVED ANALYSIS OF CARBOHYDRATES



(57) Abstract: The invention relates, in part, to the improved analysis of carbohydrates. In particular, the invention relates to the analysis of carbohydrates, such as N-glycans and O-glycans found on proteins. Improved methods, therefore, for the study of glycosylation patterns on cells, tissue and body fluids are also provided. Information regarding the analysis of glycans, such as the glycosylation patterns on cells, tissues and in body fluids, can be used in diagnostic and treatment methods as well as for facilitating the study of the effects of glycosylation/altered glycosylation on protein function. Such methods are also provided. Methods are also provided to assess protein production processes, to assess the purity of proteins produced, and to select proteins with the desired glycosylation.

WO 2005/111627 A2

**METHODS AND PRODUCTS RELATED TO THE IMPROVED ANALYSIS OF
CARBOHYDRATES**

RELATED APPLICATIONS

5 This application claims priority under 35 U.S.C. §119 from U.S. provisional application serial number 60/562,874, filed April 15, 2004, the entire contents of which is herein incorporated by reference.

GOVERNMENT SUPPORT

10 Aspects of the invention may have been made using funding from National Institutes of Health Grant number GM 57073. Accordingly, the Government may have rights in the invention.

FIELD OF THE INVENTION

15 The invention relates to the improved analysis of carbohydrates. In particular, the invention relates to the analysis of carbohydrates, such as N-glycans and O-glycans found on proteins and lipids. The invention also relates to the analysis of glycoconjugates, such as glycoproteins, glycolipids and proteoglycans. Methods for the study of glycosylation patterns on cells, tissues and in body fluid, such as serum, are also provided. Information
20 regarding the glycosylation patterns on cells can be used in diagnostic and treatment methods as well as for facilitating the study of the effects of glycosylation/altered glycosylation on diseases, protein or lipid function and function of medical treatments. Information regarding the glycosylation of glycoconjugates can also be used in the quality control analysis of glycoconjugate production and/or therapeutics.

25

BACKGROUND OF THE INVENTION

 Asparagine-linked glycosylation (N-glycosylation) is the most common co-translational modification found in eukaryotic proteins. As proteins are synthesized by the ribosome, the polypeptide enters the endoplasmic reticulum, where oligosaccharyl transferase
30 (OT) attaches a branched carbohydrate (N-glycan) to the side chain of certain asparagine residues [Hirschberg, C.B., Snider, M.D. (1987) Topography of glycosylation in the rough endoplasmic reticulum and Golgi apparatus. *Annu Rev Biochem* 56, 63-87.] This process requires an Asn-X-Ser/Thr consensus sequence in the peptide substrate, where X is any amino acid except proline [Bause, E. (1983) Structural requirements of N-glycosylation of

proteins. Studies with proline peptides as conformational probes. *Biochem J* 209, 331-6; Marshall, R.D. (1972) Glycoproteins. *Annu Rev Biochem* 41, 673-702.] After the attachment of the glycans, the carbohydrate moiety is extensively modified by a complex array of glycosidases and glycosyl transferases in the ER and golgi apparatus. The attached N-glycans are very important in protein folding, as well as directing the protein to the appropriate location within the cell [Dwek, R.A. (1996) Glycobiology: Toward Understanding the Function of Sugars. *Chem Rev* 96, 683-720; O'Connor, S.E., Imperiali, B. (1996) Modulation of protein structure and function by asparagine-linked glycosylation. *Chem Biol* 3, 803-12.] Outside the cell, the sugars aid in protein-protein interactions, often modulating the activity of the protein to which they are attached. Depending on the glycan composition, they can also protect against or facilitate protein degradation in circulation, as well as target the protein to a specific organ [Crocker, P.R., Varki, A. (2001) Siglecs in the immune system. *Immunology* 103, 137-45; Helenius, A., Aebi, M. (2001) Intracellular functions of N-linked glycans. *Science* 291, 2364-9; Imperiali, B., O'Connor, S.E. (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol* 3, 643-9.]

N-glycans also have an essential role in normal biology, as evidenced by the high lethality in cases of defective glycosylation. In mouse knockout models, disrupting even one of the biosynthetic enzymes can lead to enormous multisystemic disorders, and several result in embryonic lethality [Furukawa, K., Takamiya, K., Okada, M., Inoue, M., Fukumoto, S. (2001) Novel functions of complex carbohydrates elucidated by the mutant mice of glycosyltransferase genes. *Biochim Biophys Acta* 1525, 1-12.] There are currently six recognized human congenital disorders of glycosylation (CDGs), all resulting in patients with multiple organ abnormalities, developmental delay and immune problems, among others [Jaeken, J., Matthijs, G. (2001) Congenital disorders of glycosylation. *Annu Rev Genomics Hum Genet* 2, 129-51; Freeze, H.H., Aebi, M. (1999) Molecular basis of carbohydrate-deficient glycoprotein syndromes type I with normal phosphomannomutase activity. *Biochim Biophys Acta* 1455, 167-78; Carchon, H., Van Schaftingen, E., Matthijs, G., Jaeken, J. (1999) Carbohydrate-deficient glycoprotein syndrome type IA (phosphomannomutase-deficiency). *Biochim Biophys Acta* 1455, 155-65.] In fact, the immune system is one of the most commonly studied systems where N-glycans have been shown to play an important physiological role. For example, specific carbohydrate structures are recognized by selectins, a family of proteins expressed on endothelial cells or lymphocytes that can trigger the

immune system upon activation [Powell, L.D., Sgroi, D., Sjoberg, E.R., Stamenkovic, I., Varki, A. (1993) Natural ligands of the B cell adhesion molecule CD22 beta carry N-linked oligosaccharides with alpha-2,6-linked sialic acids that are required for recognition. *J Biol Chem* 268, 7019-27; Sgroi, D., Varki, A., Braesch-Andersen, S., Stamenkovic, I. (1993)

5 CD22, a B cell-specific immunoglobulin superfamily member, is a sialic acid-binding lectin. *J Biol Chem* 268, 7011-8.] The same class of structures that are necessary for proper immune function can also provide a binding site for certain viruses, bacteria or tumor cells in the body [Karlsson, K.A. (1998) Meaning and therapeutic potential of microbial recognition of host glycoconjugates. *Mol Microbiol* 29, 1-11; Pritchett, T.J., Brossmer, R., Rose, U., Paulson, 10 J.C. (1987) Recognition of monovalent sialosides by influenza virus H3 hemagglutinin. *Virology* 160, 502-6.]

Viral infection is mediated by the interaction of viral proteins with N-glycans on the cell surfaces of the host [Van Eijk, M., White, M.R., Batenburg, J.J., Vaandrager, A.B., Van Golde, L.M., Haagsman, H.P., Hartshorn, K.L. (2003) Interactions of Influenza A virus with 15 Sialic Acids present on Porcine Surfactant Protein D. *Am J Respir Cell Mol Biol.*] Despite the increasing evidence associating glycans to different pathogenic conditions, in multiple instances it is unclear whether changes in N-glycan structure are a cause or a symptom of the disorder. In cystic fibrosis, increased antennary fucosylation (α 1-3 linked to GlcNAc) is observed on surface membrane glycoproteins of airway epithelial cells [Glick, M.C., 20 Kothari, V.A., Liu, A., Stoykova, L.I., Scanlin, T.F. (2001) Activity of fucosyltransferases and altered glycosylation in cystic fibrosis airway epithelial cells. *Biochimie* 83, 743-7; Scanlin, T.F., Glick, M.C. (2000) Terminal glycosylation and disease: influence on cancer and cystic fibrosis. *Glycoconj J* 17, 617-26.]

There have also been many reports of alterations in N-glycan composition on cancer 25 cell proteins. For example, there are indications that prostate cancer cells produce prostate specific antigen (PSA) with more glycan branching than non-cancer cells [Peracaula, R., Tabares, G., Royle, L., Harvey, D.J., Dwek, R.A., Rudd, P.M., de Llorens, R. (2003) Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology* 13, 457-70; Belanger, A., van Halbeek, H., Graves, 30 H.C., Grandbois, K., Stamey, T.A., Huang, L., Poppe, I., Labrie, F. (1995) Molecular mass and carbohydrate structure of prostate specific antigen: studies for establishment of an international PSA standard. *Prostate* 27, 187-97; Prakash, S., Robbins, P.W. (2000) Glycotyping of prostate specific antigen. *Glycobiology* 10, 173-6.] Melanoma and bladder

- 4 -

cancer cells produce proteins with highly branched glycans due to an overexpression of the biosynthetic enzyme β 1,6-*N*-acetyl-glucosaminyltransferase V (GnT-V) [Chakraborty, A.K., Pawelek, J., Ikeda, Y., Miyoshi, E., Kolesnikova, N., Funasaka, Y., Ichihashi, M., Taniguchi, N. (2001) Fusion hybrids with macrophage and melanoma cells up-regulate N-acetylglucosaminyltransferase V, beta1-6 branching, and metastasis. *Cell Growth Differ* 12, 623-30; Przybylo, M., Hoja-Lukowicz, D., Litynska, A., Laidler, P. (2002) Different glycosylation of cadherins from human bladder non-malignant and cancer cell lines. *Cancer Cell Int* 2, 6.] Increased sialylation and additional branching have also been observed in cells from human breast and colon neoplasia [Lin, S., Kemmner, W., Grigull, S., Schlag, P.M. (2002) Cell surface alpha 2,6 sialylation affects adhesion of breast carcinoma cells. *Exp Cell Res* 276, 101-10; Nemoto-Sasaki, Y., Mitsuki, M., Morimoto-Tomita, M., Maeda, A., Tsuiji, M., Irimura, T. (2001) Correlation between the sialylation of cell surface Thomsen-Friedenreich antigen and the metastatic potential of colon carcinoma cells in a mouse model. *Glycoconj J* 18, 895-906; Dennis, J.W., Granovsky, M., Warren, C.E. (1999) Glycoprotein glycosylation and cancer progression. *Biochim Biophys Acta* 1473, 21-34; Fernandes, B., Sagman, U., Auger, M., Demetrio, M., Dennis, J.W. (1991) Beta 1-6 branched oligosaccharides as a marker of tumor progression in human breast and colon neoplasia. *Cancer Res* 51, 718-23.]

SUMMARY OF THE INVENTION

This invention provides, in part, methods related to the analysis of carbohydrates. In particular, the invention relates to the analysis of carbohydrates, such as N-glycans and O-glycans found on proteins and lipids. The invention also relates to the analysis of glycoconjugates, such as glycoproteins, glycolipids and proteoglycans.

In an aspect of the invention, methods of analyzing a sample containing glycoconjugates are provided. The methods include (a) separating the carbohydrates (e.g., glycans) from the sample containing the glycoconjugates, (b) determining the glycosylation site and/or occupancy of the glycoconjugates, (c) analyzing the carbohydrates (e.g., glycans) for characterization and/or quantification, and (d) determining the glycoforms and/or glycan profile of the glycoconjugates in the sample with the results obtained from steps (b) and (c) with a computational approach. In certain embodiments, the methods also include determining the occupancy of each glycan at each glycosylation site.

- 5 -

In an embodiment of the invention, step (a) includes denaturing the glycoconjugates.. In a second embodiment, the glycoconjugates are denatured with a denaturing agent. In another embodiment, the denaturing agent is detergent, urea, guanidium hydrochloride or heat. In a further embodiment, the glycoconjugates are reduced following their denaturation.

5 In yet another embodiment, the glycoconjugates are reduced with a reducing agent. The reducing agent in certain embodiments is DTT, β -mercaptoethanol or TCEP. In a further embodiment, the glycoconjugates are alkylated with an alkylating agent following their reduction. The alkylating agent in certain embodiments is iodoacetic acid or iodoacetamide.

In certain embodiments of the foregoing methods, the step of determining the
10 glycosylation sites and/or glycosylation site occupancy includes analyzing the glycoconjugates, preferably with 2D-NMR. In one embodiment, the step of determining the glycosylation site and glycosylation site occupancy includes cleaving the peptide backbone of the glycoconjugates (and optionally analyzing the cleaved fragments), cleaving and labeling
15 with a first label the glycoconjugates at their glycosylation sites of a first portion of the sample, cleaving the glycoconjugates at their glycosylation sites of a second portion of the sample, analyzing the first and second portions of the sample of glycoconjugates, and quantifying the results. The analysis of the first and second portions of the sample can be performed separately or mixed in any ratio.

In one embodiment, the glycoconjugates of the first portion are labeled with a label.
20 Preferably the label is an isotope of C, N, H, S or O. More preferably, the label is O^{18} . In another embodiment, the glycoconjugates of the second portion are unlabeled. In a further embodiment, the glycoconjugates of the second portion are labeled. In yet another embodiment, the first and second portions of the sample of glycoconjugates are analyzed separately. In still another embodiment, the first and second portions of the sample of
25 glycoconjugates are analyzed as a mixture. In another embodiment, the glycosylation site occupancy is quantified from ratios of the masses of cleaved glycoconjugates of the first and second portions of the sample.

In a further embodiment, the first and second portions of the sample are analyzed with a mass spectrometric method. Preferably, the mass spectrometric method is LC-MS, LC-
30 MS/-MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.

In still another embodiment, the step further includes generating a list of possible glycoconjugates and/or peptides, e.g., using databases. In a second embodiment, the step further includes generating a list of possible glycans.

- 6 -

The step of analyzing the glycans includes, in certain embodiments, analyzing the glycans with a mass spectrometric method, an electrophoretic method, NMR, a chromatographic method or a combination thereof. In a further embodiment, the mass spectrometric method is LC-MS, LC-MS/MS, MALDI-MS, MALDI-TOF, TANDEM-MS or
5 FTMS. Preferably, the mass spectrometric method is a quantitative MALDI-MS or MALDI-FTMS using optimized conditions. In yet another embodiment, the MALDI-MS is MALDI-MS optimized with a mixture of 6-aza-2-thiothymine (ATT) and Nafion[®] coating. In still another embodiment, the electrophoretic method is CE-LIF.

In additional embodiments, the step further includes contacting the glycans with one
10 or more glycan-degrading enzymes. In another embodiment, the one or more glycan-degrading enzymes is sialidase, galactosidase, mannosidase, N-acetylglucosaminidase or a combination thereof. In yet another embodiment, the step of analyzing the glycans includes quantifying the glycans using calibration curves of known glycan standards. In still another embodiment, the method further includes determining a peptide sequence of the
15 glycoconjugate.

In further embodiments of the foregoing methods, low abundance species are detected due to the low detection limits, which preferably extend to lower than about 5 fmol. Low abundance species include, but are not limited to, fucoses, sialic acids, galactoses, mannoses and sulfate groups.

20 According to another aspect of the invention, methods of analyzing a sample containing glycoconjugates are provided. The methods include separating glycans from the sample containing the glycoconjugates, determining the glycosylation sites and glycosylation site occupancy of the glycoconjugates, and analyzing the glycans to characterize and/or quantify the glycans. Determining the glycosylation sites and glycosylation site occupancy
25 includes cleaving and labeling with a first label the glycoconjugates of a first portion of the sample at their glycosylation sites, cleaving the glycoconjugates of a second portion of the sample at their glycosylation sites, analyzing the first and second portions of the sample of glycoconjugates, and quantifying the results.

In an embodiment, the glycoconjugates of the first portion are labeled. In a second
30 embodiment, the glycoconjugates of the second portion are unlabeled. In another embodiment, the glycoconjugates of the second portion are labeled. In yet another embodiment, the first and second portions of the sample of glycoconjugates are analyzed with

- 7 -

a mass spectrometric method. Preferably, the mass spectrometric method is LC-MS, LC-MS/MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.

In certain embodiments, determining the glycosylation sites and glycosylation site occupancy further includes generating a list of possible glycoconjugates. In other
5 embodiments, the step of separating the glycans from the sample includes denaturing the glycoconjugates with a denaturing agent. Preferably, the glycoconjugates are reduced with a reducing agent following their denaturation. More preferably, the glycoconjugates are alkylated with an alkylating agent following their reduction.

In further embodiments, the step of analyzing the glycans includes analyzing the
10 glycans with a mass spectrometric method, an electrophoretic method, NMR, a chromatographic method or a combination thereof. The mass spectrometric method preferably is LC-MS, LC-MS/MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS. The electrophoretic method preferably is CE-LIF. In still further embodiments, the step further includes contacting the glycans with one or more glycan-degrading enzymes.
15 Preferably the one or more glycan-degrading enzymes is sialidase, galactosidase, mannosidase, N-acetylglucosaminidase or a combination thereof. In another embodiment, the method further includes determining a peptide sequence of the glycoconjugate.

According to still another aspect of the invention, methods of determining the glycosylation site occupancy of glycoconjugates in a sample are provided. The methods
20 include cleaving and labeling with a first label the glycoconjugates at their glycosylation sites of a first portion of the sample, cleaving the glycoconjugates at their glycosylation sites of a second portion of the sample, analyzing the first and second portions of the sample of glycoconjugates, and quantifying the results. In an embodiment, the method further includes determining the possible fragments of the glycoconjugate.

In other embodiments, the glycoconjugates of the first portion are labeled with an isotope of C, N, H, S or O. Preferably the label is O^{18} . In a further embodiment, the glycoconjugates of the second portion are unlabeled. In another embodiment, the glycoconjugates of the second portion are labeled. In a further embodiment, the first and second portions of the sample of glycoconjugates are analyzed with a mass spectrometric
25 method. In yet a further embodiment, the mass spectrometric method is LC-MS, LC-MS/MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.
30 method. In yet a further embodiment, the mass spectrometric method is LC-MS, LC-MS/MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.

In further embodiments of the foregoing methods, low abundance species are detected due to the low detection limits, which preferably extend to lower than about 5 fmol. Low

abundance species include, but are not limited to, fucoses, sialic acids, galactoses, mannoses and sulfate groups.

According to yet another aspect of the invention, methods of analyzing a sample containing glycans are provided. The methods include separating neutral from charged
5 glycans, and analyzing the neutral and charged glycans separately to analyze the glycan. In preferred embodiments, the analysis of the glycans is performed with MALDI-MS.

In a further aspect of the invention, methods of analyzing a glycan are provided. The methods include analyzing the glycan in the presence of Nafion[®] and 6-aza-2-thiothymine (ATT).

10 Certain embodiments of the foregoing methods are methods of analyzing the purity of a sample containing glycans. Other embodiments of the foregoing methods are methods of analyzing the glycans of a sample of a cell, a group of cells, a tissue or serum or other body fluid from a subject. Still other embodiments of the foregoing methods are high-throughput methods, in which more than one sample of glycoconjugates is analyzed. In some preferred
15 embodiments, the more than one sample of glycoconjugates are in a 96-well plate. In other preferred embodiments, the more than one sample of glycoconjugates are on a membrane.

In other embodiments of the high-throughput methods, carbohydrate cleavage is performed using enzymes such as PNGase F, endoglycosydase H, or endoglycosydase F, or chemical methods such as hydrazinolysis or alkali borohydrate cleavage. Preferably, cleavage
20 is performed in a high-throughput manner in 96-well plates or in solution. In certain other embodiments, purification is performed using solid phase extraction cartridges such as graphitized carbon columns and C-18 columns. Preferably, purification is performed in a high-throughput manner in 96-well plates. All of the foregoing steps, particularly the step of separation, can be performed with the use of robotics.

25 According to another aspect of the invention, methods of generating a glycoconjugate, (preferably glycopeptide) library are provided. The methods include cleaving the backbone of the glycoconjugate (preferably the peptides of the glycopeptides) in a sample and labeling the fragments generated with a first labeling agent, and cleaving the glycans in the sample and labeling the fragments generated in the sample with a second labeling agent. Preferably
30 the library represents all possible glycoform fragments of the sample containing the glycoconjugates. The glycoconjugates preferably are glycopeptides, glycolipids or proteoglycans.

- 9 -

In certain embodiments, the first and second labeling agent is the same labeling agent. In other embodiments, the labeling agent is an isotope of C, N, H, S or O, preferably O¹⁸. In still other embodiments, the method further includes characterizing the fragments generated from the cleavage of the glycopeptides. In some embodiments, the characterization is performed with LC-MS, LC-MS/MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS. In further embodiments, the characterizing includes characterizing the glycosylation sites, characterizing the peptides of the glycopeptides and/or characterizing the glycans.

According to a further aspect of the invention, a library of glycopeptides generated with the foregoing methods is generated.

The library can be used as an internal standard to analyze new batches of glycoconjugates by direct comparison to each labeled standard from the library. For example, the backbones of new batches of glycoconjugates can be cleaved and mixed with the labeled fragments from the library to characterize all the glycoforms present in the new batch from the ratios of labeled and unlabeled fragments. Thus, in a further aspect of the invention, methods of analyzing a sample of glycopeptides are provided. The methods include analyzing the glycopeptides, and comparing the analyzed glycopeptides with the foregoing library of glycopeptides of the foregoing embodiments. In certain embodiments, it is preferred that comparative characterization is performed using LC-MS, LC-MS/MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.

According to another aspect of the invention, methods of generating a list of glycoconjugate properties are provided. The methods include measuring two or more properties of the glycoconjugate, and recording a value for the two or more properties of the glycoconjugate to generate a list, wherein the value of the two or more properties is recorded in a computer-generated data structure. In some embodiments, one of the two or more properties of the glycoconjugates is the number of one or more types of monosaccharides of the glycoconjugate. In other embodiments, one of the two or more properties of the glycoconjugates is the total mass of the glycans of the glycoconjugate. In still other embodiments, the glycoconjugate is a glycoprotein or proteoglycan, and one of the two or more properties of the glycoconjugate is the mass of the peptide of the glycoconjugate. In yet other embodiments, the glycoconjugate is a glycolipid, and one of the two or more properties of the glycoconjugate is the mass of the lipid of the glycoconjugate. In further embodiments, one of the two or more properties of the glycoconjugate is the mass of the glycoconjugate. In

other embodiments, one of the two or more properties of the glycoconjugate is the mass of permethylated glycans.

In a further aspect of the invention, a database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more glycoconjugates is provided. The database includes one or more data units corresponding to the one or more glycoconjugates, each of the data units including an identifier that includes two or more fields, each field for storing a value corresponding to one or more properties of the glycoconjugates.

According to still another aspect of the invention, methods of analyzing the total glycome of a sample of body fluid, cells or tissues are provided. The methods include (a) analyzing all the glycans of the sample, and (b) determining a profile of the glycans of the sample. In some embodiments, the sample is optionally fractionated and/or the glycans are separated from the glycoconjugates. The cleavage, fractionation, purification and/or separation steps described elsewhere herein are optionally included in the methods.

In certain embodiments, the method further includes performing a pattern analysis on the results from (a) using computational tools. The pattern can be described as (but is not limited to) relative amounts of the components of the pattern, absolute amounts of the components of the pattern, ratios between the components of the pattern, combinations of different components of the pattern, presence or absence of any of the components of the pattern or combination of the above. In certain embodiments, the identification of the glycome pattern and the pattern analysis can be performed using computational methods. Preferably this includes an iterative process, which optionally includes one of more of the following: incorporation of all experimental data sets from the glycome analysis and other glycan characterization, generation of theoretical glycan structures, incorporation of glycan composition, incorporation of structure and property information from databases, incorporation of glycan biosynthetic pathway information, incorporation of patient (or sample origin) information such as patient history and demographics, extract features from the experimental data sets, generation of data sets with specific features, submitting the combined information to data mining analysis, establishing relationship rules and validating the patterns.

In other embodiments, step (a) includes quantifying the glycans using calibration curves of known glycan standards. In another embodiment, the method further includes recording the pattern in a computer-generated data structure. In yet another embodiment, the

method is a method for diagnostic or prognostic purposes. In a further embodiment, the method is a method for assessing the purity of the sample. In yet another embodiment, the sample is a sample of serum, plasma, blood, urine, saliva, sputum, tears, CSF, seminal fluid, feces, tissues or cells.

5 According to another aspect of the invention, methods of analysis are provided. The methods include (a) analyzing all of the glycans of a sample of body fluid, cells and/or tissues and (b) comparing the results from (a) with a known pattern. In an embodiment, the sample is a sample of serum, plasma, blood, urine, saliva, sputum, tears, CSF, seminal fluid, feces, tissues or cells.

10 In some embodiments, the methods are methods of diagnosis and the pattern is associated with a diseased state. In one preferred embodiment, the pattern associated with a diseased state is a pattern associated with cancer, such as prostate cancer, melanoma, bladder cancer, breast cancer, lymphoma, ovarian cancer, lung cancer, colorectal cancer or head and neck cancer. In other preferred embodiments, the pattern associated with a diseased state is a
15 pattern associated with an immunological disorder; a neurodegenerative disease, such as a transmissible spongiform encephalopathy, Alzheimer's disease or neuropathy; inflammation; rheumatoid arthritis; cystic fibrosis; or an infection, preferably viral or bacterial infection. In other embodiments, the method is a method of monitoring prognosis and the known pattern is associated with the prognosis of a disease. In yet another embodiment, the method is a
20 method of monitoring drug treatment and the known pattern is associated with the drug treatment. In particular, the methods (e.g., analysis of glycome profiles) are used for the selection of population-oriented drug treatments and/or in prospective studies for selection of dosing, for activity monitoring and/or for determining efficacy endpoints.

 In another aspect of the invention, methods of determining the purity of a sample are
25 provided. The methods include (a) analyzing total glycans of the sample, (b) identifying the glycan pattern of the sample, and (c) comparing the pattern with a known pattern to assess the purity of the sample. Similar methods are provided in which glycoconjugates in sample are analyzed.

 In an aspect of the invention, a method of generating the complete glycan pattern of a
30 body fluid, cells and/or tissue is provided. The method includes, (a) analyzing the glycans in a sample of the body fluid, cells and/or tissue, and (b) identifying the complete glycan pattern of the sample. In an embodiment, neutral, charged, N-linked and O-linked glycans are included in the pattern. In other embodiments, glycosaminoglycans and glycolipids are

- 12 -

included in the pattern. In a second embodiment, the sample is a sample of serum, plasma, blood, urine, saliva, sputum, tears, CSF, seminal fluid, feces, tissues or cells.

In a further aspect of the invention, methods of analyzing the total glycome of a sample are provided. The methods include determining the glycosylation site and glycosylation site occupancy of all glycoconjugates in the sample, characterizing components of the glycoconjugates and all glycans of the glycome in the sample, and matching specific glycans to glycoconjugates with a computational method.

According to another aspect of the invention, methods of analyzing a sample of glycoconjugates are provided. The methods include analyzing the glycans of the sample with an analytical method, and determining the glycoforms of the sample with a computational method. In certain embodiments of the foregoing methods, the methods include generating constraints from the experimental analysis and solving them.

A further method for matching each carbohydrate in the glycome to its glycoconjugate includes characterization of glycosylation sites and occupancy of all glycoconjugates from body fluids, determination of possible glycans at each site by comparing unlabeled, glycoconjugate fragments to labeled, deglycosylated fragments, characterization of the entire glycome from body fluids, and combination of the different datasets into the iterative computational analysis to match the glycans to the glycoconjugates.

Each of the limitations of the invention can encompass various embodiments of the invention. It is, therefore, anticipated that each of the limitations of the invention involving any one element or combinations of elements can be included in each aspect of the invention.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 shows the conserved N-glycan pentasaccharide core.

Fig. 2 illustrates classes of N-linked glycans. High-mannose structures contain up to 9 mannose residues (**Fig. 2A**). Complex type glycans are modified with hexosamines, galactoses, sialic acids and/or fucose, among other residues (**Fig. 2B**). Complex type chains can occur as mono-, bi-, tri-, and tetra-antennary structures. Also, the amount and type of sialylation differs. Hybrid structures contain characteristics of both high-mannose and complex types (**Fig. 2C**).

Fig. 3 provides the detailed pathway of N-glycan biosynthesis (<http://www.genome.ad.jp/kegg/pathway/map/map00510.html>).

Fig. 4 shows the cleavage sites of EndoH, EndoF and PNGaseF. EndoH can only act on high mannose and hybrid structures, while EndoF is effective at cleaving all classes of N-glycans. PNGaseF also cleaves all mammalian N-glycan structures.

Fig. 5 provides the MALDI-MS spectra of N-glycans from RNaseB samples prepared by various methods. Glycans after GlycoClean S (Table 2, Sample 12), with the expected high mannose peaks and significant contamination of unknown identity (**Fig. 5A**). A small amount of sample (10 µg) was prepared using a 25 mg GlycoClean H column (Table 2, Sample 17), which showed only detergent peaks (**Fig. 5B**). A larger amount of protein (50 µg) was prepared (Table 2, Sample 18), yielding the expected glycan peaks but still containing detergent contamination (**Fig. 5C**). Using a 200 mg GlycoClean H column to purify N-glycans from 150 µg of RNaseB (Table 2, Sample 20), only the high mannose saccharides were observed (**Fig. 5D**).

Fig. 6 shows the spectra from MALDI-MS of N-glycans from ovalbumin. Each labeled peak corresponds to a previously reported structure listed in Table 3.

Fig. 7 provides results from a study of N-glycans from antibody samples. **Figs. 7A** and **7B** are for samples from Applikon bioreactors, with DO=50%, pH=7 and DO=90%, pH uncontrolled, respectively. **Figs. 7C-7E** are for samples from Wave reactors. **Fig. 7C** represents the results for DO controlled, pH uncontrolled, and NaOH in the media, while **Fig. 7D** represents the results with NaHCO₃ in the media instead of NaOH. The results shown in **Fig. 7E** are for DO uncontrolled with pH=7.

Fig. 8 shows the structures and theoretical masses of N-glycans released from antibodies.

Fig. 9 MALDI-MS spectra of glycans released from serum proteins using PNGaseF and EndoF. Serum samples were treated with PNGaseF (**Fig. 9A**) or EndoF (**Fig. 9B**) and purified. While glycans were observed, the samples did not produce clean results. The peak cluster indicated by an arrow represents detergent contamination.

Fig. 10 shows a separation of neutral and acidic glycans using GlycoClean H cartridge. (a) The original mix of standards is shown in positive mode. A3 and SC1840 are additional highly charged, and do not ionize well. (b) Neutral glycans eluted off the GlycoH cartridge ionize well in positive mode, while only the charged sugars are present in (c), allowing them to be observed in negative mode. The multiple peaks in (c) arise from sodium adducts, typically one adduct per sialic acid residue.

- 14 -

Fig. 11 provides the results from MALDI-MS of *N*-glycans from human serum in neutral (left) and acidic (right) fractions. (a) and (b) represent neutral glycans prepared from two different IMPATH normal male human serum samples, while (d) and (e) show the acidic fraction. (c) and (f) are the neutral and acidic fractions of a normal human sample from
5 Biomedical Resources.

Fig. 12 provides the results of serum glycans separated by ConA. (a) SDS-PAGE of ConA flow through (Lane 2) and elution (Lane 3). vLane 1 shows molecular weight standards. vMALDI-MS of (b) neutral and (c) acidic sugars obtained from ConA elution.

Fig. 13 provides the results from protein A separation of IgG from serum. (a)
10 Glycoblots of Protein A flow-through (Lane 3) and elution (Lane 4). Lane 1 contains protein standards, while Lane 2 (negative control) contains human serum albumin (* marks where albumin would run on an SDS-PAGE gel). Only glycosylated proteins are observed in the glycoblots, so the albumin does not stain. MALDI-MS of glycans harvested from the elution fraction are shown in (b) neutral and (c) acidic. Total serum glycans are pictured in (d)
15 neutral and (e) acidic.

Fig. 14 shows the permethylation of *N*-glycans. All OH and NH groups can be permethylated. For complete reaction, it is essential that the reaction vessel is free of air and water.

Fig. 15 shows the results of MALDI-MS of permethylated glycan standards. (a)
20 Unmodified standards ionized unevenly. Permethylated standards (b) showed more uniform ionization, but generally did not have higher signal-to-noise ratios.

Fig. 16 shows the aminooxyacetyl peptide and its conjugation to *N*-glycans. The aminooxyacetate end of the synthetic peptide (top) reacts with the open form of the reducing end GlcNAc of *N*-glycans (bottom).

Fig. 17 shows the results of MALDI-MS of peptide-conjugated *N*-linked standards.
25 (a) Unmodified glycans ionize unevenly, especially charged glycans **f** and **g**. After conjugation with aminooxyacetyl peptide (b), ionization is much more uniform.

Fig. 18 shows the identification of serum *N*-glycans from MALDI-MS spectra. (a) shows neutral glycans, while (b) shows acidic glycans. Labeled peak numbers correspond to
30 entries in Table 7.

Fig. 19 shows the results of neutral *N*-glycans from PVDF digest. Only the most abundant glycans are observed.

Fig. 20 provides a MALDI spectra of glycans before (A) and after (B) applying new recipe with optimized conditions.

Fig. 21 provides results from glycan quantification using optimized matrix recipe for MALDI-MS

5 **Fig. 22** provides a schematic of an example of a methodology for analysis.

Fig. 23 provides a flowchart illustration of one example of a combined analytical-computational method for glycan analysis.

Fig. 24 provides a scheme for an exemplary method for glycoprotein analysis - glycan site occupancy analysis.

10 **Fig. 25** provides results from glycan site occupancy analysis for ribonuclease B. MS data for peptide eluting at 7.8 minutes for unlabeled sample (A) and for the 16O/18O labeled 1:1 mixture (B). The expected $[M+H]^+$ for the unlabeled peptide fragment is 476.29 Da.

Fig. 26 provides MALDI-MS spectra of *N*-glycans from RNaseB with the expected high mannose structures.

15 **Fig. 27** provides results from MALDI-MS of *N*-glycans from ovalbumin. Each labeled peak corresponds to a previously reported structure listed below.

Fig. 28 provides structures and theoretical masses of *N*-glycans released from antibodies.

Fig. 29 shows the results from an analysis of depletion of serum albumin and IgGs from serum. A) SDS gel stained with Simply blue before (lane 7 and 14) and after removal of serum albumin and IgG using different conditions (lanes 1-6 and 8-13). B) Western blot (using protein A-HRP detection) used for quantifying the removal of IgGs. Lanes 7 and 14 are without depletion and 1-6 and 8-13 are using different conditions for the removal. C) Quantification of IgG removal.

25 **Fig. 30** shows the results of protein A separation of IgG from serum. (a) Glycoblots of Protein A flow-through (Lane 3) and elution (Lane 4). Lane 1 contains protein standards, while Lane 2 (negative control) contains human serum albumin (* marks where album would run on an SDS-PAGE gel). Only glycosylated proteins are observed in the glycoblots, so the albumin does not stain.

30 **Fig. 31** shows the identification of serum *N*-glycans from MALDI-MS spectra. (a) shows neutral glycans, while (b) shows acidic glycans.

Fig. 32 provides the results from LC-MS (A) and CE-LIF(B) analysis of neutral glycome from serum.

Fig. 33 provides the MALDI-MS acidic glycome profile of saliva (A) and urine (B).

Fig. 34 provides quantitative neutral glycome profile for serum with normal (A) and low (B) IgG levels.

Fig. 35 provides alterations in serum glycomic patterns between matched healthy (A) and cancer (B) patients

Fig. 36 provides a schematic representation of an example of the computational strategy for the analysis of glycoprofile patterns.

DETAILED DESCRIPTION

It has been recognized that carbohydrates play a significant role in a variety of biological and pathological processes. However, information regarding which carbohydrates are important and how they affect biological functions is limited. Therefore, additional methods for analyzing carbohydrates are desirable. Some of the methods provided herein provide better limits of detection of glycans and/or glycoconjugates that, in some examples, can extend to lower than 5fmol.

Methods are provided herein which are directed to improved methods of analyzing carbohydrates. As used herein, the term “carbohydrate” is intended to include any of a class of aldehyde or ketone derivatives of polyhydric alcohols. Therefore, carbohydrates include starches, celluloses, gums and saccharides. Although, for illustration, the term “saccharide” or “glycan” is used below, this is not intended to be limiting. It is intended that the methods provided herein can be directed to any carbohydrate, and the use of a specific carbohydrate is not meant to be limiting to that carbohydrate only.

As used herein, the term “saccharide” refers to a polymer comprising one or more monosaccharide groups. Saccharides, therefore, include mono-, di-, tri- and polysaccharides (or glycan). Glycans can be branched or branched. Glycans can be found covalently linked to non-saccharide moieties, such as lipids or proteins (as a glycoconjugate). These covalent conjugates include glycoproteins, glycopeptides, peptidoglycans, proteoglycans, glycolipids and lipopolysaccharides. The use of any one of these terms also is not intended to be limiting as the description is provided for illustrative purposes. In addition to the glycans being found as part of a glycoconjugate, the glycans can also be in free form (i.e., separate from and not associated with another moiety). The use of the term peptide is not intended to be limiting. The method provided herein are also intended to include proteins where “peptide” is recited.

- 17 -

The methods, therefore, provided can be used to analyze glycans that are found as part of a glyconjugate or are found in free form. The methods provided are also directed to the analysis of the total glycome of a sample. The sample can be of a cell, group of cells, tissue or serum. The “total glycome” refers to all of the glycans that can found in a sample. For instance, the glycans can be in free form or they can be part of one or more glycoconjugates in the sample. The total glycome, therefore, represents all of the glycans (in free form, as part of glycoconjugates or both) in the sample. Likewise the use of the phrase “a sample of glycans” or the like is intended to include a sample containing free glycans and/or glycans as part of glycoconjugates. The sample can be, for instance, a sample of body fluid. Samples of body fluid include serum, plasma, blood, urine, saliva, sputum, tears, CSF, seminal fluid, feces, etc. The sample, can also be, as an example, a sample of a cell, group of cells or tissue.

Glycans include N- and O-glycans. For illustration, but not intended to be limiting, N-glycans are classified into three types based on their structure: high mannose, hybrid and complex [Sears, P., Wong, C.H. (1998) Enzyme action in glycoprotein synthesis. *Cell Mol Life Sci* 54, 223-52.] All N-glycans contain a conserved pentasaccharide core composed of two *N*-acetylglucosamine residues followed by three mannose saccharides (**Fig. 1**). High mannose structures contain up to six more mannoses on both branches (**Fig. 2A**), while complex structures have no additional mannoses on either arm (**Fig. 2B**). Instead, they are composed of additional hexosamines and/or galactose. Hybrid structures are mixes of both high mannose and complex structures (**Fig. 2C**). Additionally, branch termini can be capped with sialic acid (a charged monosaccharide), and the core or branches can be fucosylated. Many other rare modifications exist, including sulfate, phosphate and xylose, but these are typically not found in humans. As provided herein, the methods of analyzing glycans may include the analysis of any glycan including glycans with any of the structures described herein. The term “glycan” is also intended to include glycans that are intact (i.e., as they were originally found in a sample) or have been digested (i.e., fragment of the original glycan).

Glycans can be analyzed with a number of different methods that include different steps and different experimental techniques. The glycans can be, for example, those displayed on proteins or lipids, on the surface of cells or any of the glycans that are present in a body fluid. In general, when the glycans in a sample are part of a glycoconjugate, the sample of glycoconjugates can be first denatured with a denaturing agent.

- 18 -

A “denaturing agent” is an agent that alters the structure of a molecule, such as a protein. Denaturing agents, therefore, include agents that cause a molecule, such as a protein to unfold. Denaturing can be accomplished with any of a number of methods that are known in the art. Denaturing can be accomplished, for instance, with heat, with heat denaturation in the presence of β -mercaptoethanol and/or SDS, by reduction followed by carboxymethylation (or alkylation), etc. Reduction can be accomplished with reducing agent, such as, dithiothreitol (DTT). Carboxymethylation or alkylation can be accomplished with, for example, iodoacetic acid or iodoacetamide. Denaturation can, for example, be accomplished by reducing with DTT, β -mercaptoethanol or tri(2-carboxyethyl)phosphine (TCEP) followed by carboxymethylation with iodoacetic acid. When the glycoconjugate sample is a sample of a body fluid, such as serum, the denaturation can be accomplished with EndoF. The glycoconjugates can also be denatured with denaturing agents, such as detergent, urea or guanidium hydrochloride.

In some methods following denaturation the sample of glycoconjugates is reduced with a reducing agent. As provided above, reducing agents include DTT, β -mercaptoethanol and tri(2-carboxyethyl)phosphine (TCEP). In other methods the sample of glycoconjugates is alkylated after being reduced, such as, for example, with iodoacetic acid or iodoacetamide.

Methods of analyzing glycans of glycoconjugates can also include cleaving the glycans from the non-saccharide moiety using any chemical or enzymatic methods or combinations thereof that are known in the art. An example of a chemical method for cleaving glycans from glycoconjugates is hydrazinolysis or alkali borohydrate. Enzymatic methods include methods that are specific to N- or O-linked sugars. These enzymatic methods include the use of Endoglycosidase H (Endo H), Endoglycosidase F (EndoF), N-Glycanase F (PNGaseF) or combinations thereof. In some preferred embodiments, PNGaseF is used when the release of N-glycans is desired. When PNGaseF is used for glycan release the proteins is, for example, first unfolded prior to the use of the enzyme. The unfolding of the protein can be accomplished with any of the denaturing agents provided above.

The glycans analyzed by the methods provided herein can also be contacted with a glycan-degrading enzyme. Examples of glycan-degrading enzymes are known in the art and include sialidase, galactosidase, mannosidase, N-acetylglucosaminidase or a combination thereof. The methods provided herein also include the use of a carbohydrate-degrading enzyme. As used herein “carbohydrate-degrading enzymes” or “glycan-degrading enzymes” are enzymes that can modify a carbohydrate or glycan in some way. Some examples of

- 19 -

glycan-degrading enzymes include sialidase, galactosidase, mannosidase, N-acetylglucosaminidase or some combination thereof.

After the release of the glycan from the protein core, or when the glycans were already in free form (not part of a glycoconjugate), the sample can be purified, for instance, by precipitating the proteins with ethanol and removing the supernatant containing the glycans. Other experimental methods for removing the proteins, detergent (from a denaturing step) and salts include any methods known in the art. These methods include dialysis, chromatographic methods, etc. In one example, the purification is accomplished with a porous graphite column. In some preferred embodiments, everything but the glycans are removed from the sample. Samples can also be purified with commercially available resins and cartridges for clean-up after chemical cleavage or enzymatic digestion used to separate glycans from protein. Such resins and cartridges include ion exchange resins and purification columns, such as GlycoClean H, S, and R cartridges. Preferably, in some embodiments GlycoClean H is used for purification.

Purification can also include the removal of high abundance proteins, such as the removal of albumin and/or antibodies, from a sample containing glycans. In some methods the purification can also include the removal of unglycosylated molecules, such as unglycosylated proteins. Removal of high abundance proteins can be a desirable step for some methods, such as some high-throughput methods described elsewhere herein. In some embodiments of the methods provided, abundant proteins, such as albumin or antibodies, can be removed from the samples prior to the final composition analysis.

Prior to the analysis of a sample of glycans as provided herein the sample of glycans can also be fractionated. The sample can be fractionated so as to obtain a sample of glycans with specific subgroups of molecules. "Subgroups of molecules" include molecules of specific properties, such as charge, molecular weight, size, binding properties to other molecules or materials, acidity, basicity, pI, hydrophobicity, hydrophilicity, etc. In one embodiment the subgroup of molecules of a sample is the low abundance species, and it is the low abundance species that are analyzed with the methods. The low abundance species can contain fucoses, sialic acids, galactoses, mannoses or sulfate groups.

The fractionation can be performed using any methods known in the art. Such methods include using solid supports with immobilized proteins, organic molecules, inorganic molecules, lipids, carbohydrates, nucleic acids, etc. The fractionation can also be performed with filters, such as molecular weight cutoff filters, resins, such as cation or

- 20 -

anion exchange resins, etc. Therefore, the method provided herein can be used for the analysis of the glycans of a subgroup of molecules.

Glycans can be charged or uncharged. They can be acidic, basic or neutral. It has now been found that separately analyzing charged and uncharged glycans of a sample can provide an improvement in the analysis of glycans. Therefore, the charged and uncharged glycans can be separated prior to the analysis of the glycans, such as with an analytic method. As described further in the **Examples** provided below, such a method has now been found to clearly discriminate the glycans present in the sample. Therefore, any of the methods provided herein can include a step of separating neutral and charged glycans, such as acidic glycans. Such separation can be achieved using purification methods. For instance, in a preferred embodiment, the separation is accomplished with a porous carbon purification cartridge by eluting glycan pools with different concentrations of acetonitrile. Other methods will be known to those of skill in the art. Analysis of these separate glycan pools can then be undertaken. For instance, when using MALDI-MS, the acidic glycans can be analyzed in negative ion mode, while the neutral glycans are analyzed in positive ion mode.

In other embodiments, the glycans can be modified to improve ionization of the glycans, particularly when MALDI-MS is used for analysis. Such modifications include permethylation. An other method to increase glycan ionization is to conjugate the glycan to a peptide. Examples of the methods are described further in the **Examples** below. In other embodiments, spot methods can be employed to improve signal intensity.

Any analytic method for analyzing the glycans so as to characterize them can be performed on any sample of glycans, such analytic methods include those described herein. As used herein, to "characterize" a glycan or other molecule means to obtain data that can be used to determine its identity, structure, composition or quantity. When the term is used in reference to a glycoconjugate, it can also include determining the glycosylation sites, the glycosylation site occupancy, the identity, structure, composition or quantity of the glycan and/or non-saccharide moiety of the glycoconjugate as well as the identity and quantity of the specific glycoform. These methods include, for example, mass spectrometry, NMR (e.g., 2D-NMR), electrophoresis and chromatographic methods. Examples of mass spectrometric methods include FAB-MS, LC-MS, LC-MS/-MS, MALDI-MS, MALDI-TOF, TANDEM-MS, FTMS, etc. NMR methods can include, for example, COSY, TOCSY, NOESY. Electrophoresis can include, for example, CE-LIF. In one embodiment the glycans can be

quantified using calibration curves of known glycan standards. More details regarding examples of these methods are provide below in the **Examples**.

Other methods that can be used to analyze the saccharide composition of the glycans once released from the protein include procedures involving the labeling of the saccharides with chemical or fluorescent tags. Such methods include fluorescence assisted carbohydrate electrophoresis (FACE), HPLC or capillary electrophoresis (CE). A method for the compositional analysis of oligosaccharides using CE has been described (Rhomberg, A. J., Ernst, S., Sasisekharan, R. & Biemann, K. (1998) *Proc Natl Acad Sci U S A* 95, 4176-81).

In some embodiments, the analytic method for the characterization of the glycans includes the use of MALDI-MS. Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) techniques for the analysis of oligosaccharides have also been described (Juhasz, P. & Biemann, K. (1995) *Carbohydr Res* 270, 131-47 and Juhasz, P. & Biemann, K. (1994) *Proc Natl Acad Sci U S A* 91, 4333-7; Venkataraman, G., Shriver, Z., Raman, R. & Sasisekharan, R. (1999) *Science* 286, 537-42; Rhomberg, A. J., Shriver, Z., Biemann, K. & Sasisekharan, R. (1998) *Proc Natl Acad Sci U S A* 95, 12232-7; Ernst, S., Rhomberg, A. J., Biemann, K. & Sasisekharan, R. (1998) *Proc Natl Acad Sci U S A* 95, 4182-7; and Rhomberg, A. J., Ernst, S., Sasisekharan, R. & Biemann, K. (1998) *Proc Natl Acad Sci U S A* 95, 4176-81). Optimized MALDI-MS analytic methods are also provided herein.

Analytic methods can also comprise the use of carbohydrate- or glycan-degrading enzymes. Following enzymatic degradation the sample of degraded glycans can be further analyzed with an analytic method as described above or otherwise known in the art.

The characterization of one or more glycoconjugates with an analytic method can also include determining the identity, structure or sequence of the non-saccharide moiety of the glycoconjugate. As an example, the characterization with an analytic method can include determining the peptide (or lipid) sequence of a glycopeptide (or glycolipid). Such methods are known in the art and examples are provided in the **Examples** section below.

In some methods, such as with MALDI-MS, the matrix in which the sample of glycans is suspended may affect the quality of compositional analysis. In some embodiments the matrix preparation is caffeic acid with or without spermine. In other embodiments, the matrix preparation is DHB with or without spermine. In preferred embodiments the matrix preparation is spermine with DHB. The spermine, for example, can be in the matrix preparation at a concentration of 300 mM. The matrix preparation can also be a combination of DHB, spermine and acetonitrile. MALDI-MS can also be performed in the presence of

- 22 -

Nafion and ATT. Additionally, when using MALDI-MS to analyze the samples, instrument parameters can also be modified. These parameters may include guide wire voltage, accelerating voltage, grid values and negative versus positive mode.

The samples of glycans can be analyzed separately or they can be analyzed as a mixture. The methods provided include methods for the analysis of glycosylation of a single protein (or lipid) in a sample or a mixture of proteins (or lipids or a mixture of proteins and lipids). Such mixtures can contain glycosylated and non-glycosylated proteins and/or lipids.

A glycoconjugate, such as a glycoprotein, can exist in many glycoforms; that is, each glycosylation site may (or may not) be occupied by a specific glycan all the time. The methods provided comprise or consist of steps for determining the glycosylation site occupancy of the glycoconjugates of a sample. As used herein, the term “glycosylation site occupancy” refers to the frequency (percentage) in which one or more specific glycosylation sites on a lipid, protein or peptide is occupied by a glycan. The glycosylation site can be determined using the methods provided below in the **Examples** as well as methods that are known in the art. In one embodiment the glycosylation site occupancy is the “total glycosylation site occupancy”, which refers to the frequencies in which all of the specific glycosylation sites on a lipid, protein or peptide are occupied by a glycan. The specific glycans that occupy each specific glycosylation site can also be characterized using one or more analytic techniques.

2D-NMR provides a reliable method for the identification of N-linked and O-linked glycan site occupancy. A combination of COSY, TOCSY, NOESY experiments are first conducted on a specific quantity of a glycoprotein. Using COSY and TOCSY data, all the spin systems (amino acids) are assigned. NOESY experiments are subsequently used to determine the specific amino acid sequence. This information allows the specific identification of all the asparagines (Asn) and serine (Ser) or threonine (Thr) residues in the sample. More importantly, since NOES between the protons of the Asn, Ser or Thr side chains and proximal carbohydrate residues can be easily monitored, this allows the monitoring and quantification of site glycan occupancy at each glycosylation site. This is particularly useful for high abundance glycosylation sites.

Also provided in one aspect of the invention is a method for determining glycosylation site occupancy of a glycoconjugate. For the determination of the glycan site occupancy, such as for lower abundance glycoforms, concepts from phosphoproteomics were adapted. **Fig. 24** provides one embodiment of a method for determining glycosylation site

- 23 -

occupancy. Briefly, a well characterized batch of the glycoprotein under study is used to generate a library of labeled peptides and glycopeptides by trypsin digest. In order to order to facilitate the determination of the glycosylation sites, each glycosylated amino acid is differentially labeled. The labels that can be used include isotopes of C, N, H, S, or O. In one embodiment the glycosylated amino acids are labeled with O¹⁸ and O¹⁶ using methods known in the art. [Kaji, 2003]. After the labeling, the samples can be further analyzed. In one embodiment the glycan site occupancy is quantified from the ratios of the masses of the labeled and unlabeled fragments. In one examples, determining the glycosylation site and its occupancy can include cleaving and labeling with a first label the glycoconjugates at the glycosylation sites of a portion of the sample, cleaving the glycoconjugates at the glycosylation sites of another portion of the sample and analyzing the portion of the sample. The portions of the sample can be analyzed separately or as a mixture in any ratio. First instance, when there are two portions of the sample, the two portions can be mixed in a 1:1, 1:2, 1:3, 1:4, or 1:5 ratio. The glycosylation site occupancy method can be used to determine the identity and number of glycoforms in the sample. Therefore, a method of determining the identity and number of glycoforms in a sample comprising determining the glycosylation site occupancy of a glycoconjugate and analysis to characterize the glycoconjugates so as to determine the identity and number of glycoforms is also provided.

As illustrated in the **Examples** below, the fragment containing the partner peak with a molecular weight 2Da heavier is identified as the peptide containing the glycosylation site. By comparing the data between the glycosylated and deglycosylated samples, a preliminary identification all the peptides (or lipids when the glycoconjugate is a glycolipid) and glycopeptides (or glycolipid) are identified and a preliminary identification of the glycans is obtained. This quantitative information can be combined with a glycan analysis and used as constraints in a computational analysis, such as the one described below, to arrive at the complete characterization of the glycoprotein.

“Constraints” as used herein are one or more values or relationships to which results obtained from an analysis of a sample containing glycans can be compared to or evaluated. The constraints can, for example, be one or more mathematical equations that can be solved with the data obtained from an analysis of a sample containing glycans and/or other data obtained from other sources, such as databases or with other analytical tools. The constraints can, for example, be generated from the one or more of the results obtained from an analysis

- 24 -

of a sample of glycans and/or with other glycan/glycoconjugate information, such as the information regarding glycan synthesis or from databases.

The labels that can be used in any of the methods provided include isotopes of C, N, H, S, or O. In one embodiment the glycosylated amino acids are labeled with O¹⁸.

5 Also provided is a method of generating a library. The library consists of labeled glycoconjugates and fragments of the glycoconjugates, the fragments being the non-saccharide portions of the glycoconjugates. In one example, a library is generated by cleaving the backbone of the glycoconjugate and labeling the non-saccharide fragments and the non-saccharide portions of the glycoconjugates that result with a labeling agent. This
10 example also includes the step of cleaving the glycans from the glycoconjugate. The glycans can then be removed from the sample. The libraries so produced can be analyzed with the methods provided herein. The libraries can also be used as a standard once characterized and methods of using such libraries are also provided. In one example, a method of analyzing a sample with glycoconjugates includes cleaving the glycoconjugates, enzymatically removing
15 the glycans from the glycoconjugates and mixing the sample with a standard. The sample mixed with the standard can then be analyzed. In one embodiment, the amounts of the glycoconjugates and non-saccharide moieties of the sample and standard are compared. In one aspect of the invention the standards are also provided.

The methods provided can also comprise or consist of the steps of generating a list of
20 glycan properties. One example of such a method includes measuring 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more properties of the glycan, recording a value for the one or more properties to generate a list of the glycan properties. The method also is intended to refer to generating a list of glycoconjugate properties. A "property" as used herein is a characteristic (e.g., structural characteristic) of the glycan or glycoconjugate that provides information (e.g.,
25 structural information) about the glycan or glycoconjugate. Examples of properties include charge, chirality, nature of substituents, quantity of substituents, molecular weight, molecular length, compositional ratios of substituents or units, type of basic building blocks (saccharide, amino acid, lipid constituents), hydrophobicity, enzymatic sensitivity, hydrophilicity, secondary structure and conformation, ratio of one set of modifications to another set of
30 modifications, etc. In one embodiment the list comprises the number of one or more types of monosaccharides. The list can also include the total mass of the glycan or glycoconjugate, the mass of the non-saccharide moiety of a glycoconjugate, the mass of one and/or more modified glycans, etc. The list in one embodiment can be a data structure tangibly embodied

in a computer-readable medium, such as computer hard drive, floppy disk, CD-ROM, etc.

Table 6 represents an examples for such a data structure. The data structure of **Table 6** has a plurality of entries, where each entry encodes a value of a property. The values encoded can be encoded by any kind of value, for example, as single-bit values, single-digit hexadecimal values, or decimal values.

Therefore, also provided is a database, tangibly embodied in a computer-readable medium, wherein the database stores information descriptive of one or more glycans and/or glycoconjugates. The database comprises data units that correspond to the glycan and/or glycoconjugate. The data units include an identifier that includes one or more fields, each field storing a value corresponding to one or more properties of the glycans and/or glycoconjugates. In one embodiment the identifier includes 2, 3, 4, 5, 6, 7, 8, 9, 10 or more fields. The database, for example, can be a database of all possible glycoconjugates, glycans or can be a database of values representing a glycome profile or pattern for one or more samples. Methods of analyzing and characterizing a glycome profile or pattern is described further below.

Herein, improved methods for analyzing samples containing glycans are provided, which include a combined analytical-computational platform to achieve a thorough characterization of glycans. Therefore, any of the methods provided can be combined with computational methods. Non-limiting examples of the computational methods that can be used are illustrated in detail in the Examples. Briefly, the diverse information gathered from the different experimental techniques can be used to generate constraints. This can be done in combination with a panel of proteomics and glycomics based bioinformatics tools and databases for the efficient characterization (glycosylation site occupancy, quantification, glycan structure, etc.) of the glycan/glycoconjugate mixture of interest. The databases can be those known in the art or can be generated with the methods provided. As an example, a method of analyzing glycans with the combined analytic and computational techniques can include the steps of performing an experiment on a sample containing glycans, analyzing the results of the experiment, generating constraints and solving them. The constraints can be generated and/or solved with the data obtained from experimental results as well as other known information, such as information from databases that contain information about glycans or glycoconjugates and with other tools that analyze the properties of glycans, glycoconjugates or the non-saccharide moieties thereof, such as mass and enzyme action.

- 26 -

The constraints can be generated using, for instance, what is known of the biosynthetic pathway of glycan synthesis. Unlike DNA or protein synthesis, which are template-driven processes, glycan biosynthesis is a complex process involving a multitude of enzymes. A detailed scheme of *N*-glycan biosynthesis is shown in **Fig. 3** and the biosynthetic enzymes and their EC numbers are listed in **Table 1**. The process is initiated in the cytoplasm, with the nascent sugar attached to the ER membrane through a lipid anchor. After a glycan core of two glucosamines followed by five mannose residues is constructed, the orientation of the growing glycan is flipped to face the lumen of the ER. There, four more mannose residues are added by α -mannosyltransferase, and one branch is capped with three glucoses. At this point, oligosaccharyl transferase catalyzes the removal of the naïve glycan from its lipid anchor, and attaches it to a glycosylation site on a protein undergoing synthesis in the ER [Varki, A. (1999) *Essentials of glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.]

To ensure that the glycan can play its proper role in protein folding and transport, the three terminal glucose residues and one mannose are removed. This trimming is required for the glycan to interact with the chaperone proteins calnexin and calreticulin [Helenius, A., Aebi, M. (2001) Intracellular functions of N-linked glycans. *Science* 291, 2364-9; Parodi, A.J. (2000) Protein glucosylation and its role in protein folding. *Annu Rev Biochem* 69, 69-93.] As the correctly folded protein passes through the Golgi on its way either to secretion or the cell membrane, further glycan modifications can take place. Specifically, mannosidases can trim more mannoses off the core sugar, while a host of glycosyltransferases can add further GlcNAc, fucose, galactose, and sialic acid moieties, among others [Sears, P., Wong, C.H. (1998) Enzyme action in glycoprotein synthesis. *Cell Mol Life Sci* 54, 223-52.]

Table 1. Common enzymes involved in *N*-glycan biosynthesis

| EC # | Enzyme name | EC # | Enzyme name |
|-----------|--|-----------|---|
| 2.4.1.- | Hexosyltransferases (ALG 6, 8, 10, 11) | 2.4.1.155 | α -1,6-mannosyl-glycoprotein 6- β - <i>N</i> -acetylglucosaminyl transferase |
| 2.4.1.38 | Glycoprotein β -galactosyltransferase | 2.4.1.201 | β -1,6-mannosyl-glycoprotein 4- β - <i>N</i> -acetylglucosaminyl transferase |
| 2.4.1.68 | Glycoprotein 6- α -L-fucosyltransferase | 2.4.99.1 | β -galactoside α -2,6-sialyltransferase |
| 2.4.1.83 | Dolichyl phosphate mannose transferase | 2.5.1.- | Transferring alkyl or aryl groups, other than methyl groups |
| 2.4.1.101 | α -1,3-mannosyl-glycoprotein 2- β - <i>N</i> -acetylglucosaminyl | 2.7.1.108 | Dolichol kinase |

- 27 -

| | | | |
|-----------|---|-----------|--|
| | transferase | | |
| 2.4.1.117 | Dolichyl phosphate β -glucosyltransferase | 2.7.8.15 | Chitobiosylpyrophosphoryl dolichol synthase |
| 2.4.1.119 | Oligomannosyl transferase | 3.1.3.51 | Dolichyl-phosphatase |
| 2.4.1.130 | Oligomannosyl synthase (ALG 3, 9, 12) | 3.1.4.48 | dolichylphosphate-glucose phosphodiesterase |
| 2.4.1.132 | Glycolipid 3- α -mannosyltransferase | 3.2.1.- | Hydrolyzing <i>O</i> - and <i>S</i> -glycosyl compounds |
| 2.4.1.141 | <i>N,N'</i> -diacetylchitobiosyl pyrophosphoryldolichol synthase | 3.2.1.106 | Mannosyl-oligosaccharide glucosidase |
| 2.4.1.142 | chitobiosyldiphosphodolichol β -mannosyltransferase | 3.2.1.113 | mannosyl-oligosaccharide 1,2- α -mannosidase |
| 2.4.1.143 | α -1,6-mannosyl-glycoprotein 2- β - <i>N</i> -acetylglucosaminyl transferase | 3.2.1.114 | mannosyl-oligosaccharide 1,3-1,6- α -mannosidase |
| 2.4.1.144 | β -1,4-mannosyl-glycoprotein 4- β - <i>N</i> -acetylglucosaminyl transferase | 3.6.1.43 | Dolichol diphosphatase |
| 2.4.1.145 | α -1,3-mannosyl-glycoprotein 4- β - <i>N</i> -acetylglucosaminyl transferase | | |

The constraints are solved using mathematical and heuristic approaches known to art such as linear programming and search techniques. A more detailed illustration of one embodiment of an analytical and computational method is provided in **Fig. 23**. One of skill
5 in the art will appreciate, however, that there are a number of ways in which experimental analytic methods can be combined with computational methods to achieve the desired characterization. It is the combination which provides more efficient analysis of samples of glycans. The examples provided are not intended to be limiting in any way.

In some embodiments, the methods provided herein also include generating a list of
10 the possible compositions of glycans and their theoretical masses. The list can be based on the biosynthetic pathways for glycosylation (**Fig. 3**). An example of such a list is provided herein. The list can also be recorded in a computer-readable medium. The list can be generated with other means, such as with the results from the use of any of the methods provided herein or known in the art. In one embodiment, the method can include the use of
15 exoenzymes to cleave the glycans in order to analyze the composition of the glycans. The list can be used in any of the methods provided in order to characterize glycans. Methods that include the use of a list are also provided herein.

Protein glycosylation can affect the function of proteins or be indicative of a cause or symptom of a disease state. For many proteins, *N*- and *O*-linked glycans are an important

factor for determining proper folding, stability and resistance to degradation (which affects the half-life of the protein). In some proteins, N- and O-linked glycans play a role in the activity and/or function of the protein. In some proteins, N- and O-linked glycans are indicative of a normal or disease state. Therefore, methods are provided herein to analyze the glycosylation of a protein for a variety of reasons. The methods, therefore, provided above
5 can be used in diagnosis.

Also provided is the method described below, which can also be used for diagnostic or prognostic purposes. In this method the total profile of carbohydrates from body fluids or tissues can be examined, and in some embodiments this can be done in a high-throughput
10 format. The examination of total glycan profiles are now exceedingly accessible thanks to the recent advances in proteomic pattern diagnostics. This approach should be useful in sensing susceptible physiological alterations to the body's natural homeostasis. This method should not only serve as a fast diagnostic tool but should also help to understand the function of specific carbohydrate modifications in some diseases.

Before developing a method to study *N*-glycans from body fluids, it was important to understand the types of molecules present. For example, proteins comprise an enormous portion of serum, approximately 7% of the total wet weight [Vander, 2001]. Of this amount, over half is albumin (~50mg/ml), a protein that can be non-enzymatically glycosylated, but not *N*-glycosylated [Rohovec, 2003]. Although the overwhelming amounts of albumin can
20 obscure analysis for proteomics, it may not interfere with *N*-glycan profiling. There are also large amounts of glycosylated antibodies, which have a number of glycan structures [Bihoreau, 1997; Watt, 2003]. However, simple methods exist to separate these abundant antibodies from the less abundant glycoproteins.

When working with serum, there are several issues to consider that are not relevant
25 for single protein systems. Because the proteins in the sample are so concentrated, they can easily precipitate out of solution. Also, even though albumin does not have *N*-linked sugars, the sheer quantity present may interfere with glycan release or purification. There are several other major proteins in serum (i.e. immunoglobulins) that are *N*-glycosylated, which may overshadow the signals from less abundant proteins. However, alterations in
30 immunoglobulin glycosylation may also be correlated with changes in physiological state. To determine the contributions and/or interference from major serum proteins, several options for separating serum proteins into fractions before analysis were explored.

- 29 -

Identifying glycan structures with complex protein mixtures can be somewhat difficult. By generating a master list of all possible compositions and their theoretical masses based on biosynthetic pathways for glycosylation, all possible monosaccharide composition was assigned to each peak observed in a MALDI-MS spectrum. Such lists and databases comprising these lists are provided herein. In most cases, each mass peak corresponds uniquely to a monosaccharide assignment. However, in some instances there can be more than one potential composition. If necessary, the correct composition can be determined by coupling different separation and characterization techniques with commercially available exoenzymes that cleave the glycans only at particular linkages.

The method provided allows fast and sensitive spectrometric analysis of patterns for the total composition of glycans in body fluids such as serum, saliva, urine, tears, seminal fluid, feces, etc. Specifically the use of the analysis of these patterns can be extended for the purpose of diagnosis, prognosis and monitoring the effects of therapeutics. Using optimized methods described above, the total content of serum, saliva and urine glycome was analyzed and it was shown that specific and reproducible MALDI-MS patterns which are dependent on the source (patient) of the sample and state could be obtained. Since every signal inside the pattern corresponds to specific glycans, the alteration of these patterns are easily determined and correlated with the expression levels of the carbohydrates. These alterations can be easily determined manually or more efficiently with the help of computational analysis. Since specific alterations in these glycan patterns are associated with disease state, this method serve as reliable platform for diagnosis, prognosis and the analysis associated with therapeutics. The methods provided can also be used to profile populations to aid the development and application of patient-oriented treatments.

Methods, therefore, are provided for the determining the glycome profile of a sample. The total glycome and/or patterns deduced therefrom can be used for studying the effects of glycosylation on protein activity and/or function as in the case of glycoprotein therapeutics. Likewise, the total glycome and/or patterns deduced therefrom can also be used in methods for diagnosis, assessing prognosis and assessing drug treatment, etc.

A "glycome profile" refers to the number and kind of glycans found in a sample. The sample can contain one or more glycans and/or one or more glycoconjugates. The glycome profile can be, for example, the number and kind of a specific type of glycan (e.g., N-glycan, O-glycan, etc.). Each component of a glycome profile can correspond to a glycan or fragment thereof or a glycoconjugate or fragment thereof. The number refers to the amount

- 30 -

and can be an actual or a relative amount. The “total glycome profile” as used herein the absolute or relative number and kind of all glycans in a sample. The sample can be a sample of cells, tissue or body fluid.

To assess the glycome profile of a sample any analytic methods can be used. Some of these methods are described above; others are known in the art. For example, the analytic method can be MALDI-MS, LC-MS, LC-MS/MS, MALDI-TOF, TANDEM-MS, FTMS, NMR, HPLC, electrophoresis, capillary electrophoresis, microfluidic devices or nanofluidic devices. In a preferred embodiment the glycome profile is determined using a quantitative MALDI-MS or MALDI-FTMS in the presence of ATT and Nafion coating. To quantify the glycans, in one example, calibration curves of known glycan standards can be used.

Prior to analyzing the glycans of the sample, the sample can be fractionated. The sample can be fractionated based on properties of the glycans and/or glycoconjugates, such as but not limited to, charge, size, molecular weight, binding properties to other molecules or materials, acidity, basicity, pI, hydrophobicity and hydrophilicity. As an example, the fractionation can be performed using solid supports with immobilized proteins, organic molecules, inorganic molecules, lipids, carbohydrates, nucleic acids, etc. As a further example, the fractionation can be performed using filters, such as molecular weight cutoff filters. The fractionation can also be performed using resins, such as, cation or anion exchange resins. Any method of fractionation known in the art can be used. In one embodiment, however, the sample is not fractionated before it is analyzed.

Prior to analysis the sample, the sample can also be degraded with a chemical or enzymatic method to cleave the glycans from any glycoconjugates in the sample. Examples of enzymatic methods are provided above and include, for example, the use of PNGase F, endoglycosydase H and endoglycosydase F or combinations thereof. Chemical methods have also been described above and include hydrazinolysis or alkali borohydrate.

After chemical or enzymatic degradation the sample can then be performed in some embodiments. Purification methods were also provided above. Examples of particular purification methods include using solid phase extraction cartridges, such as graphitized carbon columns and C-18 columns.

Once a glycome profile is determined, a glycome pattern can be identified. As used herein “glycome pattern” refers to a glycome profile or subset of the profile that has been associated with a certain function (of a lipid or protein), cellular state, or pathological condition (i.e., a disease condition). A glycome pattern can be identified using a

- 31 -

computational method. An glycome pattern, like, the profile, can be represented by the relative or absolute amounts of components of the pattern or ratios between the components of the pattern. The glycome pattern can also be represented by combinations of different components or the presence or absence of a component. The pattern can also be any
5 combination of representations, such as those provided herein.

The pattern can be determined using a computational method. Examples of such computational methods are provided herein in the **Examples**. The computational method can, for example, incorporate one or more of the following to determine a glycome pattern: experimental data from analytic methods of glycome and/or glycan analysis; theoretical
10 glycan structures; glycan composition, structure, property information from databases, glycan biosynthetic pathway information, patient or sample origin information, such as patient history, demographics; extracting features from the experimental data sets, generating all possible data sets with specific features, submitting the combined information to a data mining analysis, establish the relationship rules and validating the pattern. The
15 computational method can be an iterative process. One detailed example is provided in **Fig. 3**.

The patterns that are ultimately validated can be recorded in a computer-generated data structure. A database of validated glycome patterns is, therefore, also provided herein.

The patterns can be subsequently used for, for example, diagnostic and prognostic
20 purposes and for determining the purity of a sample.

The methods provided herein include methods for determining the glycosylation of a protein and its effects on the protein's activity and/or function. The protein glycosylation can be studied with the methods provided to determine the proper folding of the protein or to determined the influence of the protein's glycosylation on the stability/and or degradation
25 resistance of the protein (indicative of the protein's half-life). Changing the composition or the degree of glycosylation of a protein can greatly influence its half-life in circulation, as well as its activity [Chang, G.D., Chen, C.J., Lin, C.Y., Chen, H.C., Chen, H. (2003) Improvement of glycosylation in insect cells with mammalian glycosyltransferases. *J Biotechnol* 102, 61-71; Perlman, S., van den Hazel, B., Christiansen, J., Gram-Nielsen, S.,
30 Jeppesen, C.B., Andersen, K.V., Halkier, T., Okkels, S., Schambye, H.T. (2003) Glycosylation of an N-terminal extension prolongs the half-life and increases the in vivo activity of follicle stimulating hormone. *J Clin Endocrinol Metab* 88, 3227-35.] For example, erythropoietin (EPO) is a glycoprotein that has been developed as a therapeutic due

to its ability to stimulate red blood cell production in the bone marrow. It has been determined that increased sialylation of EPO greatly increases its half-life in circulation[Darling, R.J., Kuchibhotla, U., Glaesner, W., Micanovic, R., Witcher, D.R., Beals, J.M. (2002) Glycosylation of erythropoietin affects receptor binding kinetics: role of electrostatic
5 interactions. *Biochemistry* 41, 14524-31.] Thus, by understanding the role of EPO glycosylation, it is possible to manufacture a more potent drug.

Similarly, methods are provided for identifying glycosylated proteins with a desired activity and/or function. In immunoglobulins, glycosylation plays an important role in the structure of the Fc region, which is important for activation of leukocytes expressing Fc
10 receptors. When glycans on the IgG Fc region are truncated, the resulting conformational changes reduce the ability of the IgG to bind to the Fc receptor[Krapp, S., Mimura, Y., Jefferis, R., Huber, R., Sondermann, P. (2003) Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J Mol Biol* 325, 979-89.] In addition, IgG glycosylation is species specific, making it essential to choose
15 the appropriate production method for protein therapeutics [Raju, T.S., Briggs, J.B., Borge, S.M., Jones, A.J. (2000) Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology* 10, 477-86.] For example, a human protein produced in a mouse cell line may not have the necessary glycans for
20 optimal function in human patients. Therefore, the immune recognition of an antibody can be assessed with the methods of analysis provided herein.

One of the major challenges during the production of glycoprotein therapeutics is to control the generation of a specific glycoform and the subsequent characterization for quality control of the product. Therefore, methods that can efficiently characterize new batches of
25 glycoprotein therapeutics are of great value to the pharmaceutical industry. For a complete characterization of glycoprotein therapeutics, information such as glycan site occupancy, carbohydrate composition and structure at each site and quantity of each carbohydrate is required.

As described below in the **Examples**, the methods for analyzing glycans found on
30 proteins, which can include antibodies, can be used to assess the quality and variability of protein production. With the recently increased focus on protein-based therapeutics by pharmaceutical companies and research laboratories, it has become important to understand how glycosylation composition is influenced by protein production methods. In the field of

- 33 -

bioprocess engineering, there are many different types of bioreactors available for protein production. Depending on the model, parameters such as pH and dissolved oxygen (DO) can be controlled in several ways, and agitation methods can result in wide variations in shear stress. In addition, the cell-feeding process during fermentation can be altered to change the cell growth profile. All of these variables can affect protein glycosylation—even using identical conditions in two different bioreactors causes changes in glycan patterns [Kunkel, J.P., Jan, D.C., Butler, M., Jamieson, J.C. (2000) Comparisons of the glycosylation of a monoclonal antibody produced under nominally identical cell culture conditions in two different bioreactors. *Biotechnol Prog* 16, 462-70; Zhang, F., Saarinen, M.A., Itle, L.J., Lang, S.C., Murhammer, D.W., Linhardt, R.J. (2002) The effect of dissolved oxygen (DO) concentration on the glycosylation of recombinant protein produced by the insect cell-baculovirus expression system. *Biotechnol Bioeng* 77, 219-24; Senger, R.S., Karim, M.N. (2003) Effect of Shear Stress on Intrinsic CHO Culture State and Glycosylation of Recombinant Tissue-Type Plasminogen Activator Protein. *Biotechnol Prog* 19, 1199-209; Muthing, J., Kemminer, S.E., Conradt, H.S., Sagi, D., Nimtz, M., Karst, U., Peter-Katalinic, J. (2003) Effects of buffering conditions and culture pH on production rates and glycosylation of clinical phase I anti-melanoma mouse IgG3 monoclonal antibody R24. *Biotechnol Bioeng* 83, 321-34.] Therefore, provided herein are methods for analyzing the glycosylation of proteins to assess protein production methods and to determine the purity or homogeneity of glycosylated proteins produced.

Therefore, methods are provided for the direct characterization of subsequent samples of the proteins under study (as in the cases of new batches of glycoprotein therapeutics). Examples of this is described herein. One example is as follows. A well characterized batch of the glycoprotein under study is used to generate a library of backbone-labeled peptides and glycopeptides by enzymatic digest using methods known in the art [Gehrmann, 2004; Yao, 2003; Reynolds, 2002; Yao, 2001]. Trypsin proteolytic digest cleavage can be employed before and after glycan cleavage in order to expand the peptide library. Peptide labeling can be performed using methods known to experts in the art. Each characterized and quantified peptide and glycopeptide can be used to generate calibration curves using LC-MS or LC-MS/MS techniques. These peptides and glycopeptides can then be mixed (in known concentrations) with the peptide/glycopeptide mixture resulting from the trypsin proteolytic cleavage digest of the new sample batch under study. The co-elution of the labeled peptides with the unknown peptides followed by the co-detection (the ratio between labeled and

- 34 -

unlabeled peptides) using mass spectrometry allows the quantification of each peptide (and therefore the different glycoforms) in the unknown sample. In addition to the peptide/glycopeptide analysis, by splitting the flow from LC column (before entering the electrospray source) to a collection plate, the respective glycans from the eluted glycopeptides can be analyzed using the methods described herein. The use of other well established methods (e.g., hydrazide column, peptide) for the determination of glycan site occupancy can also be used [Cointe, 2000, Hui, 2002, An, 2003].

The methods provided, where the amount or type of glycans on proteins can be determined, can be used to analyze the purity of a protein sample. As used herein the term “purity” refers to the proportion of a protein sample that contains a particular glycan or a particular glycosylation pattern. In some embodiments, the protein sample is determined to be at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more pure. In some embodiments the method is used to assess the amount of a particular glycan in a protein sample. In some instances, it may be desired that the proteins are selected depending on the particular glycosylation pattern they exhibit. As used herein, “glycosylation” is meant to include the pattern or a subset or even one particular glycan, while “glycosylation pattern” refers to the number and kind of glycans present on the protein. In other aspects of the invention the methods provided herein can be used to evaluate a process of producing proteins and/or compare a process with another to evaluate the types of proteins produced. The “types of proteins produced” includes not only the protein itself but also its glycosylation pattern.

As stated above, the glycosylation of a protein may be indicative of a normal or a disease state. Therefore, methods are provided for diagnostic purposes based on the analysis of the glycosylation of a protein or set of proteins, such as the total glycome. The methods provided herein can be used for the diagnosis of any disease or condition that is caused or results in changes in protein glycosylation. For example, the methods provided can be used in the diagnosis of cancer, inflammatory disease, benign prostatic hyperplasia (BPH), etc.

The diagnosis can be carried out in a person with or thought to have a disease or condition. The diagnosis can also be carried out in a person thought to be at risk for a disease or condition. “A person at risk” is one that has either a genetic predisposition to have the disease or condition or is one that has been exposed to a factor that could increase his/her risk of developing the disease or condition. In some embodiments, the person can have, be thought to have or is at risk of cancer, cystic fibrosis, mad cow disease, etc.

Detection of cancers at an early stage is crucial for its efficient treatment. Despite advances in diagnostic technologies, many cases of cancer are not diagnosed and treated until the malignant cells have invaded the surrounding tissue or metastasized throughout the body. Although current diagnostic approaches have significantly contributed to the detection of cancer, they still present problems in sensitivity and specificity.

Cancers or tumors also include but are not limited to adrenal gland cancer, biliary tract cancer; bladder cancer, brain cancer; breast cancer; cervical cancer; choriocarcinoma; colon cancer; endometrial cancer; esophageal cancer; extrahepatic bile duct cancer; gastric cancer; head and neck cancer; intraepithelial neoplasms; kidney cancer; leukemia; lymphomas; liver cancer; lung cancer (e.g. small cell and non-small cell); melanoma; multiple myeloma; neuroblastomas; oral cancer; ovarian cancer; pancreas cancer; prostate cancer; rectal cancer; sarcomas; skin cancer; small intestine cancer; testicular cancer; thyroid cancer; uterine cancer; urethral cancer and renal cancer, as well as other carcinomas and sarcomas.

Protein samples, therefore, may include samples from a subject. The samples can, for example, be serum or saliva samples.

The methods can also be used to determine whether or not cells are undergoing dramatic change or are "stressed cells". Stressed cells are cells that are undergoing a stress response that alters the cell's protein production. The stress response can be any change that causes altered protein production or causes the cell to deviate from its normal state. Stressed cells can be identified by analyzing the glycans exhibited by the proteins on the cell's surface. Such glycans can be found in, for example, a glycoprotein. In some embodiments, the methods provided are used to detect changes in glycosylation that occur under growth conditions or inflammation.

In other aspects of the invention methods for analyzing blood type antigens are also provided.

In other aspects of the invention methods are provided for therapeutics. The glycosylation of proteins can be assessed to evaluate treatment regimens and/or to select specific therapies.

A subject is any human or non-human vertebrate, e.g., dog, cat, horse, cow, pig. A sample includes any sample obtained from any of these subjects.

High-throughput methods are also provided. "High-throughput" methods refer to the ability to process and/or analyze multiple samples at one time. High-throughput methods

- 36 -

provided herein can include the use of a membrane-based method, such as a PVDF membrane in a 96-well plate, for high throughput sample processing (i.e., digestion and/or denaturation steps, etc.) In some preferred embodiments membrane based high-throughput methods may also include the removal of abundant proteins such as albumin. In one aspect of the invention, therefore, the methods of analysis provided are high-throughput methods. Any step or steps of any of the methods provided herein can be performed as a high-throughput step. For instance purification, degradation, etc. can be performed in a high-throughput manner in some embodiments.

Robotics can be used in one or more steps of the methods provided herein. In one embodiment robotics is used for separation.

The present invention is further illustrated by the following **Examples**, which in no way should be construed as further limiting. The entire contents of all of the references (including literature references, issued patents, published patent applications, and co-pending patent applications) cited throughout this application are hereby expressly incorporated by reference.

EXAMPLES

Example 1 – N-Glycan Analysis

Materials and Methods

PNGaseF digest of N-Glycans from Protein Cores

Between 10 and 100µg of protein was denatured for 10 minutes at 90°C with 0.5% SDS and 1% β-mercaptoethanol. Since SDS (and other ionic detergents) inhibits enzyme activity, 1% NP-40 was added to counteract these effects. The enzyme reaction was performed overnight with 2µl of PNGaseF at 37°C in a 50mM sodium phosphate buffer, pH 7.5.

Purification of Released N-Glycans

Proteins were precipitated with a 3X volume of 100% ethanol on ice for 1 hour. After centrifugation to remove the proteins, the supernatant containing the N-glycans was evaporated by vacuum (SpeedVac, TeleChem International, Inc., Sunnyvale, CA). Dried glycans were resuspended in 50µl of water.

- 37 -

Samples were desalted using 1ml ion exchange column of AG50W X-8 beads (Bio-Rad, Hercules, CA). The resin was charged with 150mM acetic acid and washed with water. Glycan samples were loaded onto the column in water and washed through with 3ml H₂O. This flow through was collected and lyophilized to obtain the desalted sugars.

5 GlycoClean R and S cartridges were purchased from Prozyme (San Leandro, CA; formerly Glyko). GlycoClean R cartridges were primed with 3ml of 5% acetic acid, and the samples were loaded in water. Sugars were eluted with 3ml of water passed through the column. For GlycoClean S, the membrane was primed with 1ml water and 1ml 30% acetic acid, followed by 1ml acetonitrile. The glycan sample was loaded (in a maximum volume of 10µl) onto the disc, and the glycans were allowed to adsorb for 15 minutes. After washing the disc with 1ml of 100% acetonitrile and 5 x 1ml of 96% acetonitrile, glycans were eluted with 3 x 0.5ml water.

GlycoClean H cartridges were purchased from Prozyme (200mg bed) or ThermoHypersil (25mg bed). To prepare the GlycoClean H cartridge, the column 15 (containing 200mg of matrix) was washed with 3ml of 1M NaOH, 3ml H₂O, 3ml 30% acetic acid, and 3ml H₂O to remove impurities. The matrix was primed with 3ml 50% acetonitrile with 0.1% TFA (Solvent A) followed by 3ml 5% acetonitrile with 0.1% TFA (Solvent B). After loading the sample in water, the column was washed with 3ml H₂O and 3ml Solvent B. Finally, the sugars were eluted using 4x0.5ml of Solvent A. GlycoClean H cartridges can be 20 reused after washing with 100% acetonitrile and re-priming with 3ml of Solvent A followed by 3ml of Solvent B. For the 25mg cartridge, wash volumes were reduced to 0.5ml. Eluted fractions were lyophilized and the isolated glycans were resuspended in 10-40µl H₂O.

MALDI-MS of N-Glycans

25 Several MALDI-MS matrix compounds were tested in this study. First, caffeic acid was added to 30% acetonitrile to make a saturated solution, with or without 300 mM spermine. Alternatively, a saturated solution of dihydroxybenzoic acid (DHB) in water was used with or without 300 mM spermine. To prepare the sample spots, three methods were used. For the crushed spot method, 1µl of matrix was spotted on the stainless steel MALDI- 30 MS sample plate and allowed to dry. After crushing the spot with a glass slide, 1µl of matrix mixed 1:1 with sample was spotted on the seed crystals and allowed to dry. Alternatively, 1µl of matrix was applied followed by 1µl of sample, or vice versa. All spectra were taken with the following instrument parameters: accelerating voltage 22000V, grid voltage 93%,

guide wire 0.15% and extraction delay time of 150nsec (unless otherwise noted). All *N*-glycans were detected in linear mode with delayed type extraction and positive polarity.

Results

5 With an IgG-producing mouse hybridoma cell line, the effects of DO and pH control on cell metabolism and growth kinetics using two different reactor types was investigated. It was determined whether the IgG glycan profile was altered by the different reactor conditions. For a complete glycan analysis, the procedure for glycan isolation was optimized. The purification and analysis was performed using two known *N*-glycosylated standards with
10 different properties, ribonuclease B (RNaseB), a glycoprotein that only contains high mannose structures [29], and ovalbumin, which contains both hybrid and complex glycan structures at just one glycosylation site [30]. After finding the best methods for glycan analysis, the procedure was applied to samples (Hamel laboratory, MIT Bioprocess Engineering Center, Cambridge, MA) produced under various conditions.

15 There are several required steps for *N*-glycan analysis from proteins. While it is possible to study both the intact glycoprotein and glycopeptides from digested proteins, these types of analysis make it difficult to determine the exact composition of the glycan structures. Therefore, the intact glycan was removed from the core protein. Then, the sugar structures were separated from the protein core, purified and analyzed using methods that can provide
20 specific saccharide compositions in an accurate manner.

Release and Purification of N-Glycans from Protein Standards

There are several methods, both enzymatic and chemical, to separate glycans from their protein cores. Of the chemical methods, hydrazinolysis provides the most efficient
25 release of glycans [31]. However, both *N*- and *O*-linked glycans are released using this method, and must be separated afterwards. The sample must be very clean, with no residual salts, and the reaction does not proceed efficiently in air or water, making hydrazinolysis somewhat undesirable as a quick measure of quality control.

Several enzymatic methods are available that are specific to *N*-linked sugars.
30 Endoglycosidases H and F (EndoH and EndoF) cleave between the two interior GlcNAc residues of the glycan core, while protein *N*-glycanase F (PNGaseF) cleaves between the interior GlcNAc and the asparagine side chain of the protein core [32-34].

EndoH only acts on high mannose or hybrid structures, while EndoF can cleave complex glycans. With EndoH and EndoF, information about fucosylation on the reducing end GlcNAc is lost since this residue remains attached to the protein core. On the other hand, PNGaseF releases the entire glycans and can cleave all classes of *N*-glycans, making it a tool of choice for *N*-glycan release.

For optimal enzyme activity, proteins should be unfolded prior to digestion with PNGaseF. Typically, a protein sample can be denatured by heating in the presence of β -mercaptoethanol and/or SDS. After PNGaseF cleavage, samples contain a mixture of free glycans, protein, detergent (from the denaturing step), and salts. In some instance it is preferred that everything except the glycans are removed from the sample. To achieve this, the proteins were first precipitated with ethanol and the supernatant containing the glycans was then dried under vacuum (SpeedVac) and resuspended in water. At this point, the most difficult component to get rid of was the detergent, which interferes with some types of analytical techniques.

There are several commercially available resins and cartridges for *N*-glycan clean-up after PNGaseF digest. In addition to an ion exchange resin (AG50W X-8 from Bio-Rad), three types of purification columns from Prozyme (formerly Glyko) were tested—the GlycoClean H, S and R cartridges (Glyco H, S and R). Glyco R contains a reverse phase material that allows glycans to flow through, while retaining peptides and detergents. Glyco S is a small membrane that adsorbs the sugars in >90% acetonitrile, while hydrophobic molecules are washed away. The glycans can then be eluted with water. Glyco H, on the other hand, is a porous graphitic carbon matrix which retains both neutral and charged sugars, while allowing salts to be washed away with a low concentration of acetonitrile. The sugars can then be eluted with higher acetonitrile concentrations. Proteins and detergents typically remain on the Glyco H column. Overall, the Glyco H cartridge yielded the best results in these studies (**Table 2**).

Glycan Analysis by MALDI-MS

Numerous analytical techniques have been applied to study *N*-glycans, including mass spectrometry, NMR, electrophoresis, and chromatographic methods. NMR, for instance, can provide detailed structural information in a single experiment. Due to the lack of natural chromophores in *N*-linked carbohydrates, many of the procedures require the labeling of saccharides with chemical tags or fluorescent labels to facilitate detection. In fluorescence

assisted carbohydrate electrophoresis (FACE), glycans are fluorescently labeled and run on a polyacrylamide gel [35]. Glycan bands can then be excised for further structural analysis. Similar methods use HPLC or capillary electrophoresis (CE) for greater sensitivity and better separation. However, these techniques merely yield migration times of a sample's
5 components, giving limited structural information.

One of the simplest and most sensitive glycan analysis methods is MALDI-MS, which has detection limits in the femto- to picomole range. In addition, many samples can be analyzed in a single experiment within minutes. MALDI-MS is a soft ionization technique that utilizes an organic matrix to absorb and transfer the ionizing energy from the laser. This
10 technique is useful for many applications, from small molecules to large proteins over 100 kDa. However, sample ionization is sensitive to instrument conditions as well as sample preparation.

In particular, the matrix used to suspend the sample is important for good ionization. The efficiency of a particular matrix can vary widely, depending on the nature of the sample.
15 Multiple matrix preparations were tested, namely caffeic acid (saturated solution in 30% acetonitrile) with or without spermine, and DHB (saturated solution in water) with or without spermine. In addition, several spotting methods were evaluated: spotting 1 μ l of sample followed by 1 μ l of matrix, spotting matrix followed by sample, or mixing the two before spotting. Whether using the crushed spot method to promote matrix crystallization would
20 improve signal intensity [36] was also investigated. When acquiring the MALDI-MS data, the data collection was optimized by varying instrument parameters such as guide wire voltage, accelerating voltage, grid values, as well as negative vs. positive mode.

To evaluate the MALDI-MS conditions and calibrate the masses, commercially available *N*-glycan standards (NGA2 and NGA3) were used. In addition, we used RNaseB
25 and ovalbumin as model glycoproteins to determine the effects of sample preparation on spectra quality and to optimize glycan release.

MALDI conditions for *N*-glycan analysis were optimized using the matrix and sample preparation conditions shown in **Table 2**. Among the matrix preparations used, DHB with spermine displayed the best results. Typically, spermine is used to allow glycans to be
30 detected in negative mode, but it enhanced the glycan signals even in positive mode. The neutral glycans had poor signals in negative mode. Using the crushed spot method did not make a significant difference in signal intensity.

Table 2. Optimization of conditions for MALDI-MS and *N*-glycan clean-up.

| Sample | Matrix | Sample info | Results and comments |
|---------------------|---|---|--|
| 1. NGA2, NGA3 | DHB, saturated solution in H ₂ O, 300mM spermine | 1µl matrix on plate, add 1µl sample | Signal okay. Not too much noise |
| 2. NGA2, NGA3 | “ ” | 1µl sample on plate, add 1µl matrix | Better signal than Sample 1 or 3. This method used in all subsequent samples unless otherwise noted. |
| 3. NGA2, NGA3 | “ ” | Mix sample and matrix, spot 1µl. | Lower signal than Sample 1 |
| 4. NGA3 | Caffeic acid, 30% ACN, saturated solution | | Good signal intensity but significant unidentified adduct |
| 5. NGA3 | Caffeic acid, 30% ACN, 300mM spermine | | Large unidentified contamination peak |
| 6. NGA3 | Caffeic acid, 30% ACN | Crushed spot method | Good signal intensity but also more noise than Sample 2 |
| 7. NGA3 | DHB, saturated solution in H ₂ O | | Low signal intensity compared to Sample 2 or 5 |
| 8. NGA3 | DHB, saturated solution in H ₂ O | Acc voltage 18000, guide wire 0.1%. | Comparable to Sample 7 |
| 9. NGA3 | DHB, saturated solution in H ₂ O, 300mM spermine | | Good signal |
| 10. NGA3 | “ ” | Acc voltage 18000, guide wire 0.1%. | |
| 11. NGA3 | “ ” | Negative mode | Very low signal, almost undetectable |
| 12. 500µg RNaseB | “ ” | GlycoS | Some high mannose peaks, many unidentified peaks |
| 13. 500µg RNaseB | “ ” | AG50W X-8 Column and batch mode | Both spots spread a lot, no signal |
| 14. 500µg RNaseB | “ ” | Glyco R | Spot does not dry properly |

| | | | |
|------------------------------------|-----|--------------------|--|
| 15. 500µg RNaseB | “ ” | Glyco H (200mg) | Good signal, Man-5 through Man-9 |
| 16. 500µg Ovalbumin | “ ” | Glyco H (200mg) | Good signal, 30 peaks that match published reports |
| 17. 10µg RNaseB or ovalbumin | “ ” | Glyco H (25mg) | Spots spread, mostly contamination peaks in 1000-1300 Da range. Probably detergent. |
| 18. 50µg RNaseB | “ ” | Glyco H (25mg) | Spot spreads a lot, significant detergent contamination peaks. |
| 19. 15µg RNaseB | “ ” | Glyco H (200mg) | Good signal, very slight contamination that does not interfere with signal. GlycoH column used for future experiments. |
| 20. 150µg RNaseB | “ ” | Glyco H (200mg) | Good signal, very clean |

Using commercially available glycans and known protein standards, it was determined that the optimal method for purifying glycans after PNGaseF release was to use GlycoClean H cartridges containing 200mg of the stationary material. This method allowed
5 resulted in MALDI-MS spectra of *N*-glycans from RNaseB and ovalbumin that were consistent with published reports. The ion exchange resin did not remove all the detergents from the sample, causing the sample spots to spread on the MALDI-MS target and not crystallize properly. GlycoClean R, on the other hand, removed detergents but did not completely remove salt, which subsequently interfered with matrix crystallization and spectra
10 quality. GlycoClean S yielded acceptable sample spots on the MALDI-MS target, but failed to remove all contamination.

Fig. 5 shows spectra of some of the representative RNaseB samples from **Table 2**, with glycans purified under different conditions. In the cleanest samples, all glycan masses correspond with high mannose structures (Man-5 through Man-9).

15 To validate the reproducibility of the method, ovalbumin was used as a protein standard with complex type *N*-glycans. Optimized purification and MALDI-MS conditions were used (Glyco H 200mg, DHB matrix with spermine). The MALDI-MS data displayed results comparable to previously published reports [30]. **Fig. 6** shows the MALDI-MS spectrum of ovalbumin glycans, and **Table 3** lists the observed peaks and their structures.

20 **Table 3.** *N*-glycan structures from ovalbumin (Harvey et al, 2000)

| Peak | Structure | Theoretical Mass |
|------|-----------|------------------|
| 1 | | 1136.4 |
| 2 | | 1298.5 |
| 3a | | 1339.5 |
| 3b | | 1339.5 |
| 4 | | 1460.5 |
| 5 | | 1501.5 |
| 6a | | 1542.6 |
| 6b | | 1542.6 |
| 7 | | 1663.6 |
| 8 | | 1704.6 |

[illegible]

| | | |
|-----|--|--------|
| 13 | | 2028.7 |
| 14a | | 2069.7 |
| 14b | | 2069.7 |
| 15a | | 2110.8 |
| 15b | | 2110.8 |
| 16 | | 2151.8 |
| 17 | | 2272.8 |

- 46 -

| | | |
|----|--|--------|
| 18 | | 2313.9 |
| 19 | | 2475.9 |
| 20 | | 2638.0 |

MALDI-MS Analysis Of N-Glycans from Antibodies Produced in Applikon and Wave Reactors

Two antibody samples produced by mouse-mouse hybridoma cells (Biokit SA,
 5 Barcelona, Spain) grown in an Applikon stirred tank reactor (STR) were analyzed, along with
 three samples produced in Wave reactors. The reactor conditions used are shown in **Table 4**.

Table 4. Reactor conditions used to produce antibody samples.

| Sample | Reactor Type | DO | pH | Other |
|--------|--------------|----------------|----------------|--------------------------------------|
| 1 | Applikon STR | 50% | 7 | |
| 2 | Applikon STR | 90% | Not controlled | |
| 3 | Wave | Controlled | Not controlled | |
| 4 | Wave | Controlled | 7 | NaHCO ₃ for pH control |
| 5 | Wave | Not controlled | 7 | Fresh media for pH control |

10

In the Applikon STR reactor, pH can be controlled automatically by the instrument,
 which dispenses CO₂, NaHCO₃ and O₂ as needed. In the Wave reactor, however,

- 47 -

measurements must be taken manually and pH adjusted by hand. The pH in this reactor can be controlled by either adding fresh media as the cells grow, or adding NaHCO_3 for increased buffering capacity, and CO_2 as needed. The main difference between the reactor types is the mode of agitation. In the Applikon STR, a blade stirrer keeps the cell suspension in motion, while a sparger introduces oxygen to the system in a controlled manner. In the Wave reactor, a rocking motion generates waves that mix the components of the system and aids the transfer of oxygen and other gases into the system.

The purified antibodies were processed according to the optimized method described above. For each sample, 100 μg of protein was used as the starting material. Both positive and negative ion modes were used in the MALDI-MS to determine whether there were charged sugars present. No signal was observed in the negative mode, indicating that only neutral sugars were obtained from the antibodies. The positive ion mode MALDI-MS data of the five antibody samples are shown in **Fig. 7**. Glycoproteins produced using different conditions are shown in **Table 4**. All fractions contained the same six glycans at 1317 Da, 1463 Da, 1478 Da, 1625 Da, 1641 Da and 1787 Da. The structures corresponding to these peaks are shown in **Fig. 8** with their theoretical masses.

These results indicate that the production method did not significantly alter the occurrence of the glycans; rather, the ratios between glycans seemed to be affected. Notably, samples prepared in the Wave reactor had a lower amount of the 1625.4 Da glycan with respect to the other glycans, as well as significant reductions in the relative peak heights at 1640.9 and 1787.7 Da. Altering the culture conditions within a reactor type did not affect the relative abundance of the *N*-glycans.

While the exact mechanisms for producing these changes are not known, it is interesting that the largest changes occurred due to reactor type, not reactor conditions such as pH, DO or media composition. In previous studies, pH above 7.2 was shown to affect glycosylation composition [28]. However, for the two samples in this study with pH uncontrolled, the pH was between 6.8 and 7.2 throughout the culture period. Studies of DO effects on glycosylation demonstrated the largest differences at extremes (10% or 190%) [26], while the samples studied here were produced under moderate DO conditions (between 50% and 90%). Because the Applikon STR and the Wave reactors differ most in their method of agitation, reactor configuration is therefore the most likely source of glycan variation.

- 48 -

Differences in protein glycosylation have been linked to shear stress, as can be generated by the stirring blade or the gas sparger in an STR reactor. However, the turbulence created in the Wave reactor also generates shear stress. One hypothesis for the shear stress effect is that cells must increase their overall protein production in response to membrane and/or cytoskeletal damage. As a consequence, the biosynthetic enzymes for glycosylation are diverted away from the protein of interest [27].

Although most observed parameters, including total antibody production, were similar in Applikon STR and Wave cultures, cells from the Wave reactor had slight increases in metabolic rates. Changes in cell metabolism may yield effects similar to those caused by shear stress, as all glycoproteins synthesized in the cell must compete for the same machinery in the ER and golgi.

Example 1 References

1. Hirschberg, C.B., Snider, M.D. (1987) Topography of glycosylation in the rough endoplasmic reticulum and Golgi apparatus. *Annu Rev Biochem* **56**, 63-87.
2. Bause, E. (1983) Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem J* **209**, 331-6.
3. Marshall, R.D. (1972) Glycoproteins. *Annu Rev Biochem* **41**, 673-702.
4. Dwek, R.A. (1996) Glycobiology: Toward Understanding the Function of Sugars. *Chem Rev* **96**, 683-720.
5. O'Connor, S.E., Imperiali, B. (1996) Modulation of protein structure and function by asparagine-linked glycosylation. *Chem Biol* **3**, 803-12.
6. Crocker, P.R., Varki, A. (2001) Siglecs in the immune system. *Immunology* **103**, 137-45.
7. Helenius, A., Aebi, M. (2001) Intracellular functions of N-linked glycans. *Science* **291**, 2364-9.
8. Imperiali, B., O'Connor, S.E. (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol* **3**, 643-9.
9. Furukawa, K., Takamiya, K., Okada, M., Inoue, M., Fukumoto, S. (2001) Novel functions of complex carbohydrates elucidated by the mutant mice of glycosyltransferase genes. *Biochim Biophys Acta* **1525**, 1-12.
10. Jaeken, J., Matthijs, G. (2001) Congenital disorders of glycosylation. *Annu Rev Genomics Hum Genet* **2**, 129-51.

11. Freeze, H.H., Aeby, M. (1999) Molecular basis of carbohydrate-deficient glycoprotein syndromes type I with normal phosphomannomutase activity. *Biochim Biophys Acta* **1455**, 167-78.
12. Carchon, H., Van Schaftingen, E., Matthijs, G., Jaeken, J. (1999) Carbohydrate-
5 deficient glycoprotein syndrome type IA (phosphomannomutase-deficiency). *Biochim Biophys Acta* **1455**, 155-65.
13. Powell, L.D., Sgroi, D., Sjoberg, E.R., Stamenkovic, I., Varki, A. (1993) Natural ligands of the B cell adhesion molecule CD22 beta carry N-linked oligosaccharides with alpha-2,6-linked sialic acids that are required for recognition. *J Biol Chem* **268**,
10 7019-27.
14. Sgroi, D., Varki, A., Braesch-Andersen, S., Stamenkovic, I. (1993) CD22, a B cell-specific immunoglobulin superfamily member, is a sialic acid-binding lectin. *J Biol Chem* **268**, 7011-8.
15. Karlsson, K.A. (1998) Meaning and therapeutic potential of microbial recognition of
15 host glycoconjugates. *Mol Microbiol* **29**, 1-11.
16. Pritchett, T.J., Brossmer, R., Rose, U., Paulson, J.C. (1987) Recognition of monovalent sialosides by influenza virus H3 hemagglutinin. *Virology* **160**, 502-6.
17. Sears, P., Wong, C.H. (1998) Enzyme action in glycoprotein synthesis. *Cell Mol Life Sci* **54**, 223-52.
- 20 18. Varki, A. (1999) *Essentials of glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Parodi, A.J. (2000) Protein glucosylation and its role in protein folding. *Annu Rev Biochem* **69**, 69-93.
20. Chang, G.D., Chen, C.J., Lin, C.Y., Chen, H.C., Chen, H. (2003) Improvement of
25 glycosylation in insect cells with mammalian glycosyltransferases. *J Biotechnol* **102**, 61-71.
21. Perlman, S., van den Hazel, B., Christiansen, J., Gram-Nielsen, S., Jeppesen, C.B., Andersen, K.V., Halkier, T., Okkels, S., Schambye, H.T. (2003) Glycosylation of an N-terminal extension prolongs the half-life and increases the in vivo activity of
30 follicle stimulating hormone. *J Clin Endocrinol Metab* **88**, 3227-35.
22. Darling, R.J., Kuchibhotla, U., Glaesner, W., Micanovic, R., Witcher, D.R., Beals, J.M. (2002) Glycosylation of erythropoietin affects receptor binding kinetics: role of electrostatic interactions. *Biochemistry* **41**, 14524-31.

23. Krapp, S., Mimura, Y., Jefferis, R., Huber, R., Sondermann, P. (2003) Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J Mol Biol* **325**, 979-89.
24. Raju, T.S., Briggs, J.B., Borge, S.M., Jones, A.J. (2000) Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology* **10**, 477-86.
25. Kunkel, J.P., Jan, D.C., Butler, M., Jamieson, J.C. (2000) Comparisons of the glycosylation of a monoclonal antibody produced under nominally identical cell culture conditions in two different bioreactors. *Biotechnol Prog* **16**, 462-70.
26. Zhang, F., Saarinen, M.A., Itle, L.J., Lang, S.C., Murhammer, D.W., Linhardt, R.J. (2002) The effect of dissolved oxygen (DO) concentration on the glycosylation of recombinant protein produced by the insect cell-baculovirus expression system. *Biotechnol Bioeng* **77**, 219-24.
27. Senger, R.S., Karim, M.N. (2003) Effect of Shear Stress on Intrinsic CHO Culture State and Glycosylation of Recombinant Tissue-Type Plasminogen Activator Protein. *Biotechnol Prog* **19**, 1199-209.
28. Muthing, J., Kemminer, S.E., Conradt, H.S., Sagi, D., Nimtz, M., Karst, U., Peter-Katalinic, J. (2003) Effects of buffering conditions and culture pH on production rates and glycosylation of clinical phase I anti-melanoma mouse IgG3 monoclonal antibody R24. *Biotechnol Bioeng* **83**, 321-34.
29. Joao, H.C., Dwek, R.A. (1993) Effects of glycosylation on protein structure and dynamics in ribonuclease B and some of its individual glycoforms. *Eur J Biochem* **218**, 239-44.
30. Harvey, D.J., Wing, D.R., Kuster, B., Wilson, I.B. (2000) Composition of N-linked carbohydrates from ovalbumin and co-purified glycoproteins. *J Am Soc Mass Spectrom* **11**, 564-71.
31. Patel, T., Bruce, J., Merry, A., Bigge, C., Wormald, M., Jaques, A., Parekh, R. (1993) Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins. *Biochemistry* **32**, 679-93.
32. Tarentino, A.L., Plummer, T.H., Jr., Maley, F. (1974) The release of intact oligosaccharides from specific glycoproteins by endo-beta-N-acetylglucosaminidase H. *J Biol Chem* **249**, 818-24.

33. Tarentino, A.L., Maley, F. (1974) Purification and properties of an endo-beta-N-acetylglucosaminidase from *Streptomyces griseus*. *J Biol Chem* **249**, 811-7.
34. Tarentino, A.L., Gomez, C.M., Plummer, T.H., Jr. (1985) Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry* **24**, 4665-71.
- 5 35. Hu, G.F. (1995) Fluorophore-assisted carbohydrate electrophoresis technology and applications. *J Chromatogr A* **705**, 89-103.
36. Rhomberg, A.J., Ernst, S., Sasisekharan, R., Biemann, K. (1998) Mass spectrometric and capillary electrophoretic investigation of the enzymatic degradation of heparin-like glycosaminoglycans. *Proc Natl Acad Sci U S A* **95**, 4176-81.

10

Example 2-Profiling of N-glycans from Human Serum

Materials and Methods

Cleavage of N-Glycans from Serum Glycoproteins (Reduction/Carboxymethylation Method)

Human male normal serum samples were obtained from IMPATH (Franklin, MA) and Biomedical Resources (Hatboro, PA), and stored at -85°C. For each experiment, 50µl of serum was used to harvest N-glycans. Serum samples were first diluted 1:4 with water, then

20 DTT was added to a final concentration of 80mM. After incubation for 30 minutes at 37°C, iodoacetic acid was added to a final concentration of 400mM and incubated for 1 hour more at 37°C. The sample was dialyzed against 10mM Tris acetate pH 8.3 overnight and concentrated to ~200µl in a spin column with a 3000Da MWCO filter. To cleave the sugars from the protein, 5µl (1,000U) of PNGaseF (New England Biolabs, Beverly, MA) was added

25 and allowed to react overnight at 37°C.

Purification of N-Glycans

After glycans were cleaved from the protein, the sample was dialyzed against water to remove excess salts and glycerol (from PNGaseF formulation). Samples were then spun for

30 5 minutes at 6000xg to remove most proteins and the supernatant lyophilized to <500µl. A C18 cartridge (Waters Corporation, Milford, MA) was primed with 3ml methanol, then 3ml water, and 3ml 5% acetonitrile with 0.1% TFA. The supernatant from the spun down sample

- 52 -

was applied to the cartridge, and 3ml of 5% acetonitrile with 0.1% TFA was added to elute the glycans, while unwanted proteins were retained on the column.

GlycoClean H cartridges (Prozyme, formerly Glyko) were first primed with 3ml 1M NaOH, 3ml H₂O, 3ml 30% acetic acid, and 3ml H₂O to clean the column of any impurities.

5 Then the cartridges were washed with 3ml Solution A (50% acetonitrile, 0.1% TFA), 6ml Solution B (5% acetonitrile, 0.1% TFA), and 3ml of water. Glycan samples were loaded in a minimal volume of water (<100µl), and the column was washed with 3ml H₂O followed by 3ml Solution B. Neutral glycans were eluted with 3ml of 15% acetonitrile, 0.1% TFA, and acidic glycans were eluted with 3ml of Solution A. For the trials with glycan standards, the
10 six glycans listed in **Table 5** were mixed in equimolar amounts (1µl of 100µM each), and the mixture was applied to the GlycoClean H cartridge and processed as described above. Each fraction was dried, then redissolved in 40µl H₂O for MALDI-MS analysis. All MALDI-MS spectra were calibrated using the six glycan standards in **Table 5**. Separate calibration files were used for positive and negative modes.

Fractionation of Serum Proteins

Concanavalin A-agarose beads were purchased from Vector Laboratories (Burlingame, CA). To prepare the column, 3ml ConA-agarose slurry was washed with ConA buffer (20mM Tris, 1mM MgCl₂, 1mM CaCl₂, 500mM NaCl, pH 7.4). Before loading, 500µl
20 of serum was mixed with 150µl of 5X ConA buffer. After washing with 3ml ConA buffer, glycoproteins were eluted with 2ml of 500mM α-methyl-mannoside and dialyzed against 10mM phosphate pH 7.2 overnight at 4°C.

Protein A-agarose beads were purchased from Calbiochem (La Jolla, CA). Before use, 1ml beads were washed 3X with PBS. To separate IgG from other serum proteins,
25 samples were diluted 1:4 with PBS and incubated with protein A-agarose overnight at 4°C. Non-IgGs were collected by loading the slurry into a column and washing with 2ml PBS. IgGs were eluted with 2ml of 0.2M glycine pH 2.5 and neutralized in 200µl Tris-HCl, pH 6.3.

SDS-PAGE and Glycoblotting of Serum Samples

30 Protein samples were prepared for SDS-PAGE by diluting 1:1 with 2X denaturing buffer (40µg/ml SDS, 20% glycerol, 30µg/ml DTT and 10µg/ml bromophenol blue in

- 53 -

125mM Tris, pH 6.8), and boiling for 2min. Pre-cast Nu-PAGE 10% Bis-Tris protein gels were obtained from Invitrogen (Carlsbad, CA). Each lane was loaded with a maximum of 10µl of sample, and run for 50min at 200V. After electrophoresis was complete, the gel was stained with Invitrogen SafeStain (1 hour in staining solution, then washed overnight with water).

The GlycoTrack glycoprotein detection kit was obtained from Prozyme (formerly Glyko). All reagents except buffers were supplied with the kit. Two methods were attempted—either biotinylating glycoproteins after blotting (a) or before blotting (b). For both methods, samples were first diluted 1:1 with 200mM sodium acetate buffer, pH 5.5. The membrane was blocked by incubating overnight at 4°C with blocking reagent, then washed 3x10 minutes with TBS.

For method (a), samples were denatured with SDS sample buffer, and subjected to SDS-PAGE and blotting to nitrocellulose. After washing the membrane with PBS, the proteins were oxidized with 10ml of 10mM sodium periodate in the dark at room temperature for 20 minutes. The membrane was washed 3 times with PBS, and 2µl of biotin-hydrazide reagent was added in 10ml of 100mM sodium acetate, 2mg/ml EDTA for 60 minutes at room temperature. After 3 washes with TBS, the membrane was blocked overnight at 4°C with blocking reagent. Before adding 5µl of streptavidin-alkaline phosphatase (S-AP) conjugate, the membrane was washed again with TBS. The S-AP was allowed to incubate for 60 minutes at room temperature, and excess was washed off with TBS. To develop the blot, 50µl of nitro blue tetrazolium (50mg/ml) and 37.5µl of 5-bromo-4-chloro-3-indolyl phosphate *p*-toluidine (50mg/ml) were added in 10ml TBS, 10mg/ml MgCl₂. After 60 minutes, the blot was washed with distilled water and allowed to air dry.

In method (b), 20µl of sample was mixed with 10µl of 10mM periodate in 100mM sodium acetate, 2mg/ml EDTA and incubated in the dark at room temperature for 20 minutes. To destroy excess periodate 10µl of a 12.5mg/ml sodium bisulfite solution in 200mM NaOAc, pH 5.5 was added for 5 minutes at room temperature. Biotinylation was performed by adding 5µl of biotin amidocaproyl hydrazide solution in DMF. After incubating at room temperature for 60 minutes, the sample was mixed with SDS denaturing buffer and boiled for 2 minutes. Samples were run on SDS-PAGE gels as described above, then transferred to a nitrocellulose membrane (2hrs, 30V). At this point, blocking and developing steps were identical to method (a).

- 54 -

Chemical Modification of N-Glycans

For permethylation, glycans in water were placed in a round-bottomed flask and lyophilized overnight. A slurry of NaOH in DMSO (0.5ml) was added to the glycan sample, along with 0.5ml methyl iodide and incubated for 15 minutes. The sample was then diluted with water and extracted 2X with CHCl₃, collecting the organic phase. After drying the organic phase with MgSO₄, it was filtered through glass wool and dried under vacuum. Samples were then redissolved in methanol for MALDI-MS analysis.

To conjugate *N*-glycans to synthetic aminooxyacetyl peptide, glycans were dried and resuspended in aqueous peptide solution (240μM). After adding 1μl of 500mM NaOAc pH 5.5 and 20μl of acetonitrile, the sample was incubated overnight at 40°C. Before MALDI-MS analysis, glycopeptides were purified by C18, 0.6μl bed ZipTip (Millipore, Billerica, MA). Specifically, the tip was washed with 5μl of 100% acetonitrile, followed by water and 5% acetonitrile, 0.1% TFA. To load the sample, 2μl of sample was drawn into the tip, and discarded after 5 seconds. After washing 3X with 5μl H₂O, glycopeptides were eluted with 10% acetonitrile.

PNGaseF Digestion on PVDF Membrane

PVDF-coated wells in a 96-well plate were washed with 200μl MeOH, 3x 200μl H₂O and 200μl RCM buffer (8M urea, 360mM Tris, 3.2mM EDTA, pH 8.3). The protein samples (50μl) were then loaded in the wells along with 300μl RCM buffer. After washing with wells two times with fresh RCM buffer, 500μl of 0.1M DTT in RCM buffer was added for 1hr at 37°C. To remove the excess DTT, the wells were washed three times with H₂O. For the carboxymethylation, 500μl of 0.1M iodoacetic acid in RCM buffer was added for 30 minutes at 37°C in the dark. The wells were washed again with water, then the membrane was blocked with 1ml polyvinylpyrrolidone (360,000 AMW, 1% solution in H₂O) for 1hr at room temperature. Before adding the PNGaseF, the wells were washed again with water. To release the glycans, 4μl of PNGaseF was added in 300μl of 50mM Tris, pH 7.5 and incubated overnight at 37°C. Released glycans were pipetted from the wells and purified by C18 and GlycoClean H as described above.

Results

Building on the work with single protein systems, the purpose of this study was to isolate and purify *N*-glycans from human serum, generating a total *N*-glycan profile. Serum was chosen as the diagnostic medium because many disease markers are released into circulation[16, 17], and obtaining serum is a relatively simple procedure.

5 Before being able to develop a method to study *N*-glycans from serum, it was important to understand the types of molecules present. Proteins comprise an enormous portion of serum, approximately 7% of the total wet weight [18]. Of this amount, over half is albumin (~50mg/ml), a protein that can be non-enzymatically glycosylated, but not *N*-glycosylated [19]. Although the overwhelming amounts of albumin can obscure analysis for
10 proteomics, it may not interfere with *N*-glycan profiling. There are also large amounts of glycosylated antibodies, which have a number of glycan structures [20, 21]. However, simple methods exist to separate these abundant antibodies from the less abundant glycoproteins.

When working with serum, there are several issues to consider that are not relevant
15 for single protein systems. Because the proteins in the sample are so concentrated, they can easily precipitate out of solution. Also, even though albumin does not have *N*-linked sugars, the sheer quantity present may interfere with glycan release or purification. There are several other major proteins in serum (i.e. immunoglobulins) that are *N*-glycosylated, which may overshadow the signals from less abundant proteins. However, alterations in
20 immunoglobulin glycosylation may also be correlated with changes in physiological state. To determine the contributions and/or interference from major serum proteins, several options for separating serum proteins into fractions before analysis were explored.

There are both neutral and charged sugars on serum glycoproteins. Acidic glycans generally do not ionize well in the positive ion mode of MALDI-MS, and also suffer loss of
25 sialic acids. On the other hand, neutral sugars ionize extremely poorly in negative mode, which is commonly used for charged glycans. Therefore, a method where the neutral and acidic structures were assayed separately was compared to two chemical modification methods that allow the glycans to ionize more uniformly.

Identifying glycan structures with complex protein mixtures can be rather difficult.
30 By generating a master list of all possible compositions and their theoretical masses based on biosynthetic pathways for glycosylation, all possible monosaccharide composition was assigned to each peak observed in a MALDI-MS spectrum. In most cases, each mass peak corresponds uniquely to a monosaccharide assignment. However, in some instances there

can be more than one potential composition. If necessary, the correct composition can be determined by using commercially available exoenzymes that cleave the glycans only at particular linkages.

5 *Sample preparation*

Serum samples generally contained upwards of 120mg/ml of protein, making heat denaturing less ideal. Even when diluted, the proteins in these samples precipitated rapidly, giving the sample a gel-like consistency. This could prevent the PNGaseF from accessing all the *N*-glycan sites on the proteins. One set of samples was processed using the traditional
10 heat-denaturation method after diluting the serum samples 1:10 in water (**Fig. 9A**). A number of glycan peaks were observed in the MALDI-MS spectrum, but there clearly was residual detergent contamination from the denaturing step. On a separate sample, EndoF was used since the enzyme can act on folded proteins (**Fig. 9B**).

However, as discussed above, EndoF cleaves between the first and second GlcNAc on
15 the glycan core, causing a loss of information on core fucosylation. After EndoF digestion the samples were purified as usual. As shown in **Fig. 9**, glycans could indeed be obtained using both methods. EndoF spectra had a relatively high level of baseline noise, and signal intensities were relatively low (~1000), leaving room for improvement.

As an alternative to heat denaturation, the proteins were reduced with dithiothreitol
20 (DTT) followed by carboxymethylation with iodoacetic acid to denature the proteins [22]. Reduction disrupts the disulfide bonds in proteins, while carboxymethylation prevents the proteins from re-folding. After dialysis to remove excess iodoacetic acid and DTT, PNGaseF was added to the denatured proteins for overnight cleavage. An additional advantage to this method over the regular SDS/ β -mercaptoethanol heat-denaturation method was that the
25 absence of detergents facilitated purification. After the glycans were cleaved from the core protein, the sample was dialyzed against water overnight and lyophilized. Exchanging the sample into water prepared it to be passed through a C18 cartridge (Waters Corporation) to remove remaining protein. At this stage, both neutral and acidic sugars were present in the same sample, potentially complicating the assignment of glycan peaks in the MALDI-MS
30 spectrum.

Because serum contains proteins with a wide variety of neutral and charged *N*-glycans, analysis was facilitated by separating the neutral sugars from the acidic carbohydrates. This allowed each pool to be analyzed using methods particularly suited to

- 57 -

the chemical properties of neutral vs. charged molecules. The GlycoClean H purification cartridge (Prozyme) was used for this purpose by eluting glycan pools with different concentrations of acetonitrile. Neutral sugars were eluted with 15% acetonitrile, 0.1% TFA, while acidic sugars were eluted with 50% acetonitrile, 0.1% TFA. To test the separation of neutral and acidic sugars, six known glycan standards were used (**Table 5**).

Table 5. Glycan standards used to test GlycoClean H separation of neutral and acidic sugars.

| Commercial name | Structure description | Mass | Charge state (# sialic acids) |
|-----------------|---|--------|-------------------------------|
| NGA2 | Asialo, agalacto biantennary | 1317.2 | 0 |
| NA2 | Asialo, galactosylated, biantennary | 1641.5 | 0 |
| NA3 | Asialo, galactosylated, triantennary | 2006.0 | 0 |
| SC1223 | Disialylated, galactosylated, fucosylated biantennary | 2370.2 | 2 |
| A3 | Trisialylated, galactosylated, triantennary | 2879.9 | 3 |
| SC1840 | Tetrasialylated, galactosylated, tetrantennary | 3683.4 | 4 |

The neutral sugars were analyzed in positive ion mode in the MALDI-MS, while acidic sugars were examined in negative mode (**Fig. 10**). To confirm that no neutral sugars were present in the acidic glycan sample, the positive mode spectrum was checked for charged glycans. When this method was applied to human serum N-glycans, the spectra appeared much cleaner than those obtained from a mixed sample, since each group of sugars could be analyzed under optimal conditions (**Fig. 11**).

This process was repeated multiple times with the same serum sample to ensure reproducibility (three aliquots were purified in parallel on one day, and another two on two different days). In addition, multiple normal serum samples were processed by this method to determine the degree of glycan variation between serum samples. Five normal male human samples from each of two different sources (IMPATH tissue bank and Biomedical Resources) were used to assess whether observed glycan profiles were consistent across suppliers. As expected, there was some variation in the spectra from different normal samples in both the neutral and acidic fraction (**Fig. 11**), while aliquots from the same serum sample appeared very similar even when they were purified on different days. The samples from different serum banks showed similar profiles and major peak clusters.

Most Abundant Proteins

Although serum samples can be analyzed with all proteins present, including non-glycosylated species, it was determined whether better results could be obtained by removing proteins such as albumin. Concanavalin A (ConA) is a lectin that binds to α -linked mannose, as contained in all *N*-glycans [23]. A serum sample was passed through a column of agarose-bound ConA. Proteins containing *N*-glycans bound to the column while non-glycosylated proteins were washed off (this sample was collected as the ConA flow through). The glycoproteins were then eluted with a 500mM α -methyl-mannoside solution, which competes for the ConA binding sites.

To evaluate the separation of the serum sample into glycosylated and non-glycosylated proteins, the ConA flow through and elution samples were run on an SDS-PAGE gel (**Fig. 12A**). In the gel, the albumin fraction is clearly visible in the flow through from the ConA column, while multiple bands in the elution lane represent glycoproteins. In addition, the glycan profiles of both the flow through and the elution fraction were analyzed. After dialyzing the samples against 10mM phosphate buffer, they were processed with PNGaseF and purified by C18 cartridge and GlycoH as described above. There were no observable glycans present in the flow-through fraction, while neutral and acidic sugars from the elution fraction are shown in **Figs. 12B and 12C**. The results from total serum digests, however, yield MALDI-MS data with signal-to-noise ratio and signal intensity that are as good as or better than from ConA elution. Therefore, in some cases there will be little to no advantage to removing non-glycosylated proteins before analysis.

Serum samples were also depleted of antibodies through a Protein A column to determine how many major peaks in the final spectra came from IgG. The presence of glycoproteins in both the flow through and elution fractions were determined by GlycoTrack glycoprotein detection kit (Prozyme/Glyko) (**Fig. 13A**). The Protein A elution fraction containing IgGs was treated with PNGaseF and purified as described above. **Figs. 13B-13E** show a comparison of the glycans from IgG (Protein A elution) to the total glycan profile. Although several of the major peaks in the spectra indeed come from this antibody population, they do not appear for these samples to be in large enough quantities to interfere with the signals from other glycans.

MALDI-MS Analysis of Serum N-Glycans

Neutral and acidic sugars require different treatment when being analyzed by MALDI-MS. In particular, neutral sugars ionize well in the positive ion mode, but not well in negative mode, while the opposite is true for charged sugars. Three different matrix formulations were tested to determine the best one for these samples. All formulations contained DHB and spermine, as this had yielded the best results with single-protein studies. The three matrix preparations were 1) saturated DHB in water with 300mM spermine, 2) 20mg/ml DHB in acetonitrile and 25mM spermine in water in a 1:1 ratio and 3) 20mg/ml DHB in methanol and 25mM spermine in water in a 1:1 ratio. Preparation 2 yielded MALDI-MS spectra with the highest signal-to-noise ratio in both positive and negative mode, and was used for all experiments.

There are several reported methods for increasing the sensitivity and ionization efficiency of mass spectrometry data in the analysis of glycans. With these methods, it is sometimes possible to analyze glycan pools as a mixture of neutral and acidic glycans, as the chemical properties of the glycans are modified to allow for more uniform ionization. Two types of chemical modifications were tested to determine whether the MALDI-MS results could be improved upon.

N-glycan samples are commonly permethylated to protect each OH and NH₂ or amide group in the carbohydrate [24]. This is particularly useful for MS techniques such as fast atom bombardment (FAB-MS), since permethylated glycans fragment in a much more predictable manner than underivatized glycans. Permethylation can also increase sensitivity in electrospray (ES-MS) and MALDI-MS. The schematic of the permethylation reaction is shown in **Fig. 14**. Some drawbacks to permethylation are that the sample has to be extremely clean for the reaction to go to completion, and the sample requires clean-up after the reaction. In the current study, although this method slightly improved the ionization of *N*-glycan standards in MALDI-MS over non-modified glycans, the increase in signal-to-noise ratio was not significant (**Fig. 15**).

A newer method for increasing *N*-glycan ionization, as well as allowing the glycans to ionize more uniformly across species is to conjugate it to a peptide [25]. The structure of the peptide and its glycan conjugation reaction are shown in **Fig. 16**. Before MALDI-MS, it was necessary to clean up the reaction mixture using a C18 ZipTip in order to eliminate the buffer (NaOAc) used in the reaction. The ZipTip flowthrough, water wash and 10% acetonitrile elution were all spotted on the plate. The glycopeptide conjugates (in the 10% elution fraction) were readily observed in the MALDI-MS, and neutral and acidic sugars ionized

- 60 -

more evenly in the positive mode as compared to unmodified glycans (**Fig. 17**). While the glycan-peptide conjugation reaction is simple, the free peptide is particularly unstable. Specifically, the peptide's active hydroxylamine group readily reacts with any aldehydes or ketones present, thus preventing it from conjugating to the glycans. Although the reaction with glycan standards displayed promising results, it was difficult to obtain a complete reaction with serum samples. Even after several attempts to label serum glycans with varying amounts of peptide, free glycan peaks in the spectra were observed from flow-through and water wash spots. Because there may be excess aldehydes or ketones remaining in serum samples, peptide conjugation was not used, and the samples were analyzed as separate neutral and acidic fractions.

Identifying Composition of Glycans from MALDI-MS Data

In a MALDI-MS spectrum, the main information obtained is mass of the parent ion. With just this data, it was indeed possible to deduce the monosaccharide composition of each peak (number of hexNAc, hexose, fucose and sialic acid residues). Using our knowledge of biosynthetic rules, as well as whether each glycan is charged or uncharged, we can significantly limit the number of possible structures for each mass peak observed. A spreadsheet to use as a lookup table for unknown peaks was created. In addition to unmodified masses, entries for permethylated masses were included as well as peptide-conjugated glycans, according to the following equations:

| n =HexNAc | h =hexose | f =fucose | s =sialic acid |
|----------------------------|------------------|------------------|-----------------------|
| MW-H ₂ O= 203.1 | 162.1 | 146.1 | 291.3 |

Unmodified glycans

$$\text{mass} = 203.1n + 162.1h + 146.1f + 291.3s + 18$$

Permethylated glycans

$$\text{perm} = \text{mass} + 51 + 14[3(n+h) + 2f + 5s]$$

Peptide-conjugated glycans

$$\text{peptide} = \text{mass} + 1527.1$$

Using this table, regardless of the analytical methods, MALDI-MS peaks can be associated with specific monosaccharide compositions. Sample entries from this database are shown in **Table 6**.

Table 6. Table of sample entries for identifying *N*-glycan composition from MALDI-MS data.

| HexNAc | Hexose | Fucose | Sialic Acid | mass | perm | peptide |
|--------|--------|--------|-------------|--------|--------|---------|
| 2 | 3 | 1 | 0 | 1056.6 | 1345.6 | 2583.7 |
| 2 | 4 | 0 | 0 | 1072.6 | 1375.6 | 2599.7 |
| 3 | 3 | 0 | 1 | 1404.9 | 1777.9 | 2932 |
| 4 | 3 | 0 | 1 | 1608 | 2023 | 3135.1 |
| 4 | 3 | 2 | 0 | 1608.9 | 2009.9 | 3136 |
| 5 | 3 | 1 | 0 | 1665.9 | 2080.9 | 3193 |
| 6 | 3 | 0 | 0 | 1722.9 | 2151.9 | 3250 |
| 3 | 6 | 1 | 0 | 1746 | 2203 | 3273.1 |
| 4 | 3 | 1 | 1 | 1754.1 | 2197.1 | 3281.2 |
| 4 | 3 | 0 | 2 | 1899.3 | 2384.3 | 3426.4 |
| 4 | 3 | 2 | 1 | 1900.2 | 2371.2 | 3427.3 |

5 Using the table almost all the peaks in MALDI-MS serum profiles could be identified as glycans of known composition (**Fig. 18**). Many of the unidentified peaks are ammonium or sodium adducts. The composition and mass of each labeled peak are listed in **Table 7**. A few peaks in the acidic glycan spectrum correspond to more than one composition. This is more common in the higher mass range since there are a larger number of possible

10 monosaccharide compositions. Many of the glycans observed in these spectra were also present in other serum samples; there are typically between 25-30 neutral glycans as well as 25-30 acidic glycans present in a given sample.

Table 7. Composition and mass of serum glycans observed in **Fig. 18**

| Neutral glycans (Fig. 18A) | | | | | |
|----------------------------|--------|--------|--------|-------------|--------|
| Peak | HexNAc | Hexose | Fucose | Sialic Acid | Mass |
| 1 | 2 | 3 | 1 | 0 | 1056.6 |
| 2 | 2 | 4 | 0 | 0 | 1072.6 |
| 3 | 2 | 5 | 0 | 0 | 1234.7 |
| 4 | 3 | 3 | 1 | 0 | 1259.7 |
| 5 | 3 | 4 | 0 | 0 | 1275.5 |
| 6 | 4 | 3 | 0 | 0 | 1316.7 |
| 7 | 2 | 6 | 0 | 0 | 1396.8 |
| 8 | 3 | 4 | 1 | 0 | 1421.8 |

- 62 -

| | | | | | |
|----------------------------------|----------------------|----------------------|----------------------|---------------------------|--------------------|
| 9 | 3 | 5 | 0 | 0 | 1437.8 |
| 10 | 4 | 3 | 1 | 0 | 1642.8 |
| 11 | 4 | 4 | 0 | 0 | 1478.8 |
| 12 | 5 | 3 | 0 | 0 | 1519.8 |
| 13 | 2 | 7 | 0 | 0 | 1558.9 |
| 14 | 3 | 6 | 0 | 0 | 1599.9 |
| 15 | 4 | 4 | 1 | 0 | 1624.9 |
| 16 | 4 | 5 | 0 | 0 | 1640.9 |
| 17 | 5 | 3 | 1 | 0 | 1665.9 |
| 18 | 5 | 4 | 0 | 0 | 1681.9 |
| 19 | 4 | 5 | 1 | 0 | 1787.0 |
| 20 | 4 | 6 | 0 | 0 | 1803.0 |
| 21 | 5 | 4 | 1 | 0 | 1828.0 |
| 22 | 5 | 5 | 0 | 0 | 1844.0 |
| 23 | 2 | 9 | 0 | 0 | 1883.1 |
| 24 | 5 | 5 | 1 | 0 | 1990.1 |
| 25 | 5 | 6 | 0 | 0 | 2006.1 |
| 26 | 5 | 5 | 2 | 0 | 2136.2 |
| 27 | 5 | 6 | 1 | 0 | 2152.2 |
| 28 | 5 | 7 | 0 | 0 | 2168.2 |
| Acidic glycans (Fig. 18B) | | | | | |
| <i>Peak</i> | <i>HexNAc</i> | <i>Hexose</i> | <i>Fucose</i> | <i>Sialic Acid</i> | <i>Mass</i> |
| 1 | 4 | 4 | 0 | 1 | 1770.1 |
| 2 | 3 | 6 | 0 | 1 | 1891.2 |
| 3 | 4 | 4 | 1 | 1 | 1916.2 |
| 4 | 4 | 5 | 0 | 1 | 1932.2 |
| 5 | 5 | 4 | 0 | 1 | 1973.2 |
| 6 | 3 | 7 | 0 | 1 | 2053.3 |
| 7 | 4 | 5 | 1 | 1 | 2078.3 |
| 8 | 5 | 4 | 1 | 1 | 2119.3 |
| 9 | 5 | 5 | 0 | 1 | 2135.3 |
| 10 | 4 | 5 | 0 | 2 | 2223.5 |
| 11 | 4 | 5 | 2 | 1 | 2224.4 |
| 12 | 5 | 3 | 1 | 2 | 2248.5 |
| 13 | 5 | 4 | 0 | 2 | 2264.5 |
| 14 | 5 | 5 | 1 | 1 | 2281.4 |
| 15 | 5 | 6 | 0 | 1 | 2297.4 |
| 16 | 4 | 5 | 1 | 2 | 2369.6 |

| | | | | | |
|----|---|---|---|---|--------|
| 17 | 4 | 5 | 3 | 1 | 2370.5 |
| 18 | 5 | 6 | 1 | 1 | 2443.5 |
| 19 | 5 | 5 | 1 | 2 | 2572.7 |
| 20 | 5 | 6 | 0 | 2 | 2588.7 |
| 21 | 5 | 6 | 2 | 1 | 2589.6 |
| 22 | 6 | 4 | 1 | 2 | 2613.7 |
| 23 | 5 | 6 | 1 | 2 | 2734.8 |
| 24 | 5 | 6 | 3 | 1 | 2735.7 |
| 25 | 5 | 6 | 2 | 2 | 2880.9 |
| 26 | 5 | 6 | 4 | 1 | 2881.8 |

Alternative Sample Preparation Methods

Besides performing PNGaseF digests in solution, a membrane-based method was tested with the potential for high throughput sample processing. Proteins were adsorbed onto a PVDF membrane in a 96-well plate, followed by reduction, carboxymethylation and digestion in the wells[26]. While this method works well with single glycoproteins, very few glycans were observed when serum was used. An explanation for this result is that albumin most likely saturated the membrane binding capacity, and most of the glycoproteins were washed away before the PNGaseF was added. Without removing albumin, only a few glycans were observed in the neutral fraction, and no acidic glycans were present (**Fig. 19**). Although time saved by performing the experiment in a 96-well format may be negated by the extra steps required to remove albumin, using the PVDF membrane as a platform for digestion may be extremely useful for the development of a high-throughput glycomics methodology for serum samples. All samples in this study were, however, processed with the PNGaseF digest in solution.

It has been demonstrated that a complete *N*-glycan profile from human serum proteins can be obtained. By separating glycans into neutral and acidic pools, it was possible to clearly identify glycans directly from MALDI-MS without chemical modification. In addition, it was shown that albumin and IgGs do not need to be removed from serum samples prior to analysis. With the ability to profile all glycans from serum, it becomes possible to apply bioinformatics approaches to search for patterns that define normal or disease states.

Furthermore, a glycomics approach may be even more sensitive than what can be achieved with proteomics. Even in cases where protein expression does not change, the types

of N-glycans present on these proteins can indicate a change in physiological condition. Already, proteomics technologies are being explored as diagnostic tools. Examining glycosylation patterns may enable more precise characterization of certain disease states, such as the differentiation between benign and malignant tumors. Thus, in combination with a bioinformatics platform, serum glycan profiling could advance the utility of glycomics data for the early diagnosis of currently undetectable disease states.

Example 2 References

1. Jaeken, J., Matthijs, G. (2001) Congenital disorders of glycosylation. *Annu Rev Genomics Hum Genet* **2**, 129-51.
2. Freeze, H.H., Aebi, M. (1999) Molecular basis of carbohydrate-deficient glycoprotein syndromes type I with normal phosphomannomutase activity. *Biochim Biophys Acta* **1455**, 167-78.
3. Van Eijk, M., White, M.R., Batenburg, J.J., Vaandrager, A.B., Van Golde, L.M., Haagsman, H.P., Hartshorn, K.L. (2003) Interactions of Influenza A virus with Sialic Acids present on Porcine Surfactant Protein D. *Am J Respir Cell Mol Biol*.
4. Glick, M.C., Kothari, V.A., Liu, A., Stoykova, L.I., Scanlin, T.F. (2001) Activity of fucosyltransferases and altered glycosylation in cystic fibrosis airway epithelial cells. *Biochimie* **83**, 743-7.
5. Scanlin, T.F., Glick, M.C. (2000) Terminal glycosylation and disease: influence on cancer and cystic fibrosis. *Glycoconj J* **17**, 617-26.
6. Peracaula, R., Tabares, G., Royle, L., Harvey, D.J., Dwek, R.A., Rudd, P.M., de Llorens, R. (2003) Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology* **13**, 457-70.
7. Belanger, A., van Halbeek, H., Graves, H.C., Grandbois, K., Stamey, T.A., Huang, L., Poppe, I., Labrie, F. (1995) Molecular mass and carbohydrate structure of prostate specific antigen: studies for establishment of an international PSA standard. *Prostate* **27**, 187-97.
8. Prakash, S., Robbins, P.W. (2000) Glycotyping of prostate specific antigen. *Glycobiology* **10**, 173-6.
9. Chakraborty, A.K., Pawelek, J., Ikeda, Y., Miyoshi, E., Kolesnikova, N., Funasaka, Y., Ichihashi, M., Taniguchi, N. (2001) Fusion hybrids with macrophage and

melanoma cells up-regulate N-acetylglucosaminyltransferase V, beta1-6 branching, and metastasis. *Cell Growth Differ* **12**, 623-30.

10. Przybylo, M., Hoja-Lukowicz, D., Litynska, A., Laidler, P. (2002) Different glycosylation of cadherins from human bladder non-malignant and cancer cell lines. *Cancer Cell Int* **2**, 6.
11. Lin, S., Kemmner, W., Grigull, S., Schlag, P.M. (2002) Cell surface alpha 2,6 sialylation affects adhesion of breast carcinoma cells. *Exp Cell Res* **276**, 101-10.
12. Nemoto-Sasaki, Y., Mitsuki, M., Morimoto-Tomita, M., Maeda, A., Tsuiji, M., Irimura, T. (2001) Correlation between the sialylation of cell surface Thomsen-Friedenreich antigen and the metastatic potential of colon carcinoma cells in a mouse model. *Glycoconj J* **18**, 895-906.
13. Dennis, J.W., Granovsky, M., Warren, C.E. (1999) Glycoprotein glycosylation and cancer progression. *Biochim Biophys Acta* **1473**, 21-34.
14. Fernandes, B., Sagman, U., Auger, M., Demetrio, M., Dennis, J.W. (1991) Beta 1-6 branched oligosaccharides as a marker of tumor progression in human breast and colon neoplasia. *Cancer Res* **51**, 718-23.
15. Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., Liotta, L.A. (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572-7.
- 20 16. Pujol, J.L., Quantin, X., Jacot, W., Boher, J.M., Grenier, J., Lamy, P.J. (2003) Neuroendocrine and cytokeratin serum markers as prognostic determinants of small cell lung cancer. *Lung Cancer* **39**, 131-8.
17. Gadducci, A., Cosio, S., Carpi, A., Nicolini, A., Genazzani, A.R. (2004) Serum tumor markers in the management of ovarian, endometrial and cervical cancer. *Biomed Pharmacother* **58**, 24-38.
- 25 18. Vander, A.J., Sherman, J.H., Luciano, D.S. (2001) *Human physiology : the mechanisms of body function*. McGraw-Hill, Boston, Mass.
19. Rohovec, J., Maschmeyer, T., Aime, S., Peters, J.A. (2003) The structure of the sugar residue in glycated human serum albumin and its molecular recognition by phenylboronate. *Chemistry* **9**, 2193-9.
- 30 20. Bihoreau, N., Ramon, C., Lazard, M., Schmitter, J.M. (1997) Combination of capillary electrophoresis and matrix-assisted laser desorption ionization mass

spectrometry for glycosylation analysis of a human monoclonal anti-Rhesus(D) antibody. *J Chromatogr B Biomed Sci Appl* **697**, 123-33.

21. Watt, G.M., Lund, J., Levens, M., Kolli, V.S., Jefferis, R., Boons, G.J. (2003) Site-specific glycosylation of an aglycosylated human IgG1-Fc antibody protein generates
5 neoglycoproteins with enhanced function. *Chem Biol* **10**, 807-14.
22. Lacko, A.G., Reason, A.J., Nuckolls, C., Kudchodkar, B.J., Nair, M.P., Sundarajan, G., Pritchard, P.H., Morris, H.R., Dell, A. (1998) Characterization of recombinant human plasma lecithin: cholesterol acyltransferase (LCAT): N-linked carbohydrate structures and catalytic properties. *J Lipid Res* **39**, 807-20.
- 10 23. Bryce, R.A., Hillier, I.H., Naismith, J.H. (2001) Carbohydrate-protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A. *Biophys J* **81**, 1373-88.
24. Fukuda, M., Kobata, A. (1993) *Glycobiology : a practical approach*. IRL Press at Oxford University Press, Oxford ; New York.
- 15 25. Zhao, Y., Kent, S.B., Chait, B.T. (1997) Rapid, sensitive structure analysis of oligosaccharides. *Proc Natl Acad Sci U S A* **94**, 1629-33.
26. Papac, D.I., Briggs, J.B., Chin, E.T., Jones, A.J. (1998) A high-throughput microscale method to release N-linked oligosaccharides from glycoproteins for matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis. *Glycobiology*
20 **8**, 445-54.

Example 3-Glycan Analysis

Release of glycans from proteins

25 Several methods were used to cleave the carbohydrates from proteins:

A) Glycoproteins were denatured with 0.5% SDS and 1% β -mercaptoethanol. Since SDS (and other ionic detergents) inhibits enzyme activity, 1% NP-40 was added to counteract these effects. The enzymatic cleavage was performed overnight with PNGaseF (New England Biolabs, Beverly, MA) at 37°C in sodium phosphate buffer, pH 7.5 or Tris acetate
30 buffer pH 8.3.

B) Samples were reduced with DTT followed by alkylation with either iodoacetic acid or iodoacetamide. The sample was dialyzed against phosphate buffer, pH 7.5 or Tris acetate pH 8.3 overnight and concentrated to ~200 μ l in a spin column with a 3000Da MWCO filter.

To cleave the sugars from the protein between 100 and 2,000U of PNGaseF (New England Biolabs, Beverly, MA) were used.

C) Glycoproteins were denatured using a buffer containing 8M urea, 3.2 mM EDTA and 360 mM Tris, pH 8.6 [Papac, 1998]. Reduction and carboxymethylation of the glycoproteins was then achieved using DTT and iodoacetic acid (or iodoacetamide), respectively. After removal of denaturing, reducing and alkylating reagents, *N*-glycans were selectively released from the glycoproteins by incubation with PNGase F.

D) The steps for protein denaturing, protein alkylation and glycan release were also performed with the proteins bound to a solid support [Papac, 1998]. PVDF-coated wells in a 96-well plate were washed with 200µl MeOH, 3x 200µl H₂O and 200µl RCM buffer (8M urea, 360mM Tris, 3.2mM EDTA, pH 8.3). The protein samples (10 to 50µl) were then loaded in the wells along with 300µl RCM buffer. After washing with wells two times with fresh RCM buffer, 300µl of 0.1M DTT in RCM buffer was added for 1hr at 37°C. To remove the excess DTT, the wells were washed three times with H₂O. For the carboxymethylation, 300µl of 0.1M iodoacetic acid in RCM buffer was added for 30 minutes at 37°C in the dark. The wells were washed again with water, the membrane was then blocked with 1ml polyvinylpyrrolidone (360,000 AMW, 1% solution in H₂O) for 1hr at room temperature. Before adding the PNGaseF, the wells were washed again with water. To release the glycans, 100 to 1,000U of PNGaseF were added in 300µl of 50mM Tris, pH 7.5 and incubated overnight at 37°C. Released glycans were pipetted from the wells and purified.

E) Alternatively, after the proteins were denatured, EndoH or Endo F (instead of PNGase F) was used to release the glycans.

F) Chemical methods, such as hydrazinolysis and reductive β-elimination were also used.

G) The denaturing, reduction, alkylation and glycan cleavage steps were also performed in a semi-high-throughput fashion either in solution or by binding the proteins to solid supports in plates with hydrophobic membranes [Papac, 1998].

Purification of Released N-Glycans

Several methods were used to isolate and purify the released carbohydrates. These methods were used either individually and some were used in combination.

- 68 -

A) Proteins were precipitated with a 3X volume of cold ethanol. After centrifugation to remove the proteins, the supernatant containing the *N*-glycans was evaporated by vacuum (SpeedVac, TeleChem International, Inc., Sunnyvale, CA). Dried glycans were resuspended in water.

5 B) Concomitant protein and salt removal was achieved using cation exchange column of AG50W X-8 beads (Bio-Rad, Hercules, CA). The resin was charged with 150mM acetic acid and washed with water. Glycan samples were loaded onto the column in water, and washed through with 3ml H₂O. The flow through was collected and lyophilized to obtain the desalted sugars.

10 C) GlycoClean R cartridges (Prozyme, San Leandro, CA; formerly Glyko) were primed with 3ml of 5% acetic acid, and the samples were loaded in water. Sugars were eluted with 3ml of water passed through the column.

D) GlycoClean S cartridges (Prozyme, San Leandro, CA; formerly Glyko), were primed with 1ml water and 1ml 30% acetic acid, followed by 1ml acetonitrile. The glycan
15 sample was loaded (in a maximum volume of 10μl) onto the disc, and the glycans were allowed to adsorb for 15 minutes. After washing the disc with 1ml of 100% acetonitrile and 5 x 1ml of 96% acetonitrile, glycans were eluted with 3 x 0.5ml water.

E) GlycoClean H cartridges (Prozyme; 200mg bed) were washed with 3ml of 1M NaOH, 3ml H₂O, 3ml 30% acetic acid, and 3ml H₂O to remove impurities. The matrix was
20 primed with 3ml 50% acetonitrile with 0.1% TFA (Solvent A) followed by 3ml 5% acetonitrile with 0.1% TFA (Solvent B). After loading the sample in water, the column was washed with 3ml H₂O and 3ml Solvent B. Finally, the sugars were eluted using 4x0.5ml of Solvent A. GlycoClean H cartridges can be reused after washing with 100% acetonitrile and re-priming with 3ml of Solvent A followed by 3ml of Solvent B. For the 25mg cartridge,
25 wash volumes were reduced to 0.5ml. Eluted fractions were lyophilized and the isolated glycans were resuspended in 10-40μl H₂O.

F) Hypercarb SPE cartridges (Thermo) were washed with 3ml of 1M NaOH, 3ml H₂O, 3ml 30% acetic acid, and 3ml H₂O to remove impurities. The matrix was primed with
30 3ml 5% acetonitrile with 0.05% TFA (Solvent B). After loading the sample in water, the column was washed with 3ml H₂O and 3ml Solvent B. Finally, the neutral sugars were eluted using 15% acetonitrile 0.05% TFA and acidic glycans were eluted using 50% acetonitrile 0.05% TFA.

- 69 -

G) Non-porous graphitic carbon SPE cartridges (SUPELCO) were primed with 3ml 5% acetonitrile and 0.05% TFA (Solvent B). After loading the sample in water, the column was washed with 3ml H₂O and 3ml Solvent B. Finally, the neutral sugars were eluted using 15% acetonitrile 0.05% TFA and acidic glycans were eluted using 50% acetonitrile 0.05% TFA.

H) The glycan purification step was also performed in a high-throughput format by using columns in 96-well plates. This process was facilitated by the use of a TECAN robot. This protocol allowed the processing of more than 90 samples at the same time.

10 *Chemical Modification of N-Glycans*

Several derivatization methods are currently used to increase the sensitivity and ionization efficiency of mass spectrometry data in the analysis of glycans. With these methods, it is often possible to analyze glycan pools as a mixture of neutral and acidic glycans, as the chemical properties of the glycans are modified to allow for more uniform ionization. N-glycan samples are commonly permethylated to protect each OH and NH₂ or amide group in the carbohydrate. This is particularly useful for MS techniques such as fast atom bombardment (FAB-MS), since permethylated glycans fragment in a more predictable manner than underivatized glycans. Permethylation can also increase sensitivity in electrospray (ES-MS) and MALDI-MS.

For permethylation, glycans in water were placed in a round-bottomed flask and lyophilized overnight. A slurry of NaOH in DMSO (0.5ml) was added to the glycan sample, along with 0.5ml methyl iodide and incubated for 15 minutes. The sample was then diluted with water and extracted 2X with CHCl₃, collecting the organic phase. After drying the organic phase with MgSO₄, it was filtered through glass wool and dried under vacuum. Samples were then redissolved in methanol for MALDI-MS analysis. Some drawbacks to permethylation are that the sample has to be extremely clean for the reaction to go to completion and requires additional purification after the reaction. Although this method slightly improved the ionization of N-glycan standards in MALDI-MS over unmodified glycans, many species corresponding to incomplete modification were detected.

To conjugate N-glycans to the synthetic aminooxyacetyl peptide, glycans were dried and resuspended in aqueous peptide solution (240μM). After adding 1μl of 500mM NaOAc pH 5.5 and 20μl of acetonitrile, the sample was incubated overnight at 40°C. Before MALDI-MS analysis, glycopeptides were purified by C18, 0.6μl bed ZipTip (Millipore,

- 70 -

Billerica, MA). Specifically, the tip was washed with 5 μ l of 100% acetonitrile, followed by water and 5% acetonitrile, 0.1% TFA. To load the sample, 2 μ l of sample was drawn into the tip, and discarded after 5 seconds. After washing 3X with 5 μ l H₂O, glycopeptides were eluted with 10% acetonitrile. Before MALDI-MS, it was necessary to clean up the reaction mixture using a C18 ZipTip in order to eliminate the buffer (NaOAc) used in the reaction. The ZipTip flowthrough, water wash and 10% acetonitrile elution were all spotted on the plate. The glycopeptide conjugates (in the 10% elution fraction) were readily observed in the MALDI-MS, and neutral and acidic sugars ionized more evenly in the positive mode as compared to unmodified glycans.

While the glycan-peptide conjugation reaction is simple, the free peptide is particularly unstable. Specifically, the peptide's hydroxylamine group readily reacts with any aldehydes or ketones present, thus preventing it from conjugating to the glycans. Other labeling reagents (i.e. APTS, ANTS, AMAC, etc.) have been used but the analysis of unmodified glycans, separated into neutral and acidic fractions, was the method of choice for these studies.

MALDI-MS analysis optimization of unmodified Glycans

Neutral and acidic sugars require different treatment when being analyzed by MALDI-MS. In particular, neutral sugars ionize well in the positive ion mode, while the ionization of acidic sugars is optimal using the negative ion mode. To be able to analyze the low abundance glycans present in a mixture glycoforms or different glycoproteins, a matrix of matrices containing more than 96 possible recipe combinations was generated. This study was designed to optimize the MALDI-MS analysis for the highest sensitivity, spot morphology, reduced peak splitting, reduced fragmentation and linear response as a function of concentration.

As a starting point, the matrix for glycans (DHB) was utilized in combination with spermine (20mg/ml DHB in acetonitrile and 25mM spermine in water in a 1:1 ratio.). This recipe resulted in detection limits of 1 pmol and 10 pmol for neutrals and acidic glycan respectively. Significant peak splitting with multiple sodium and potassium ions were observed. Also, this matrix crystallized as long needle-shaped crystals, which makes it difficult to achieve reproducible quantification of glycans present in a sample and eliminates the possibility for the automation of data acquisition.

- 71 -

Some of the matrices and reagents used in this study were: caffeic acid, dihydroxybenzoic acid (DHB), spermine, 1-hydroxyisoquinoline (HIQ), 6-aza-2-thiothymine (ATT), 2,4,6-trihydroxyacetophenone (THAP), Nafion, 6-hydroxypicolinic acid, 3-hydroxypicolinic, 5-methoxysalicylic acid (5-MSA), ammonium citrate, ammonium tartrate, sodium chloride, ammonium resins, etc. These reagents were used in combination with different solvents such as methanol, ethanol, acetonitrile and water. The matrix of matrices study resulted in new recipes of 2,5-dihydroxybenzoic acid (5 mg/ml) and 5-methoxysalicylic acid (0.25 mg/ml) in acetonitrile for neutrals and 6-aza-thiothymine (10 mg/ml in Ethanol) spotted on Nafion coating for acidic glycans. These matrices displayed the best detection limits for a mixture of carbohydrates to our knowledge: 25 fmol and 5 fmol for neutrals and acidic glycans respectively (**Fig. 20**). The new matrices also showed minimum peak splitting, highly uniform signal intensity, spot morphology and no detectable fragmentation.

A detailed study to correlate between signal intensity, concentration and molecular weight was also performed. The analysis covered the entire range of possible molecular weights for N-glycans (900 – 4200 Da). Linear response as a function of concentration was observed for different glycans. Taken together, MALDI analysis of glycans using these matrices can be used to quantify the amount of glycans present in a mixture (**Fig. 21**). In particular these data enable the quantification of glycans at the low fmole concentration range. Other methods known to those of ordinary skill in the relevant art can also be used to quantify glycans at a higher range of concentrations [Harvey, 1993]. For **Figs. 1 and 2**, the assigned peaks and the labels correspond to glycan standards from Dextra Laboratories Ltd. (Reading, United Kingdom).

A potential concern in MALDI is that the ion yield of specific analyte in a mixture drops as the number of constituents increase. To evaluate this, the effect of the signal strength on the number of glycans present in a mixture was also evaluated for both matrices. Interestingly, there was very little change in the intensity of individual glycan signals even in the presence of other glycans, thus indicating that the ion yield of a specific constituent is not affected by the number of analytes present in the glycan mixture (**Fig. 21B**). This ensures that even in a complex mixture of glycans accurate amounts can be calculated using the signal intensity. Finally, the dynamic range of these matrices were in the low fmol range ensuring that changes in low abundant glycans can be accurately monitored by using these matrices.

- 72 -

To prepare the sample spots, three methods were used. For the crushed spot method, 1 µl of matrix was spotted on the stainless steel MALDI-MS sample plate and allowed to dry. After crushing the spot with a glass slide, 1 µl of matrix mixed 1:1 with sample was spotted on the seed crystals and allowed to dry. Alternatively, 1 µl of matrix was applied followed by 1 µl of sample, or vice versa. When resins were used in combination with the matrices, 1 µl of the resin was applied to the probe and allowed to dry before applying the sample in a 1:1 mixture with the matrix. All spectra were taken with the following instrument parameters: accelerating voltage 22000V, grid voltage 93%, guide wire 0.15% and extraction delay time of 150 nsec (unless otherwise noted). All N-glycans were detected in linear mode with delayed extraction and positive polarity for neutrals and negative polarity for acidics.

LC-MS, LC-MS/MS and capillary electrophoresis

Due to the limitations in isomass characterization using MALDI-MS, in some instances other techniques such as LC-MS (or MS/MS) and CE-LIF can be applied to further characterize the glycans released from the glycoprotein of interest. For LC-MS (or MS/MS), the reducing end of the carbohydrates is reduced using sodium borohydride and the carbohydrates are separated in the using graphitized carbon column. The column is directly attached to an electrospray ionization mass spectrometer (ESI-MS) which allows the detection and characterization of the carbohydrates as they elute from the column. Although the use of exoglycosidases is often added to this LC-MS analysis, MS/MS fragmentation is also used for further linkage characterization of the carbohydrates based on the fragmentation pattern.

Similarly, capillary electrophoresis-laser induced fluorescence (CE-LIF) can also be used for the further separation and characterization of the glycans. In this case, the carbohydrates, are first derivatized, in some preferred embodiments, by a reductive amination, at their reducing end with a fluorescent molecule such as APTS, ANTS, AMAC, etc. The fluorescently-modified (or "labeled") carbohydrates are then separated by capillary electrophoresis and detected with high sensitivity *via* laser induced fluorescence. Similar to LC-MS, glycosidases can also be used in combination with CE-LIF in order to get further structural linkage information on the carbohydrates.

Identifying Glycans Composition from MALDI-MS Data

- 73 -

In a MALDI-MS spectrum, the main information obtained is mass of the parent ion. With just this data, it was indeed possible to deduce the monosaccharide composition of each peak (number of hexNAc, hexose, fucose and sialic acid residues). Using available information of biosynthetic rules, as well as whether each glycan is charged or uncharged, the number of possible structures for each mass peak observed can be significantly limited. A spreadsheet to use as a lookup table for unknown peaks was created. In addition to unmodified masses, entries for permethylated masses were included as well as peptide-conjugated glycans, according to the following equations:

| | |
|----|--|
| 10 | $n = \text{HexNAc}$ $h = \text{hexose}$ $f = \text{fucose}$ $s = \text{sialic acid}$ |
| | $\text{MW-H}_2\text{O} = 203.1 \quad 162.1 \quad 146.1 \quad 291.3$ |
| | Unmodified glycans |
| | $\text{mass} = 203.1n + 162.1h + 146.1f + 291.3s + 18$ |
| | Permethylated glycans |
| 15 | $\text{perm} = \text{mass} + 51 + 14[3(n+h) + 2f + 5s]$ |
| | Peptide-conjugated glycans |
| | $\text{peptide} = \text{mass} + 1527.1$ |

Using this table, regardless of the analytical methods, mass spectrometry peaks can be associated with specific monosaccharide compositions. A table of sample entries was shown above in **Table 6**. Other methods known to the art can be used to determine the glycan identity from MS data (See, for example U.S. Pat. No. 5,607,859; U.S. Ser. No. 09/558, 137; and WO 00/65521).

Computational tools to characterize glycoprotein mixture

The diverse information gathered from the different experimental techniques are incorporated as constraints and used in combination with a panel of proteomics and glycomics based bioinformatics tools and databases for the efficient characterization (glycosylation site occupancy, quantification, glycan structure, etc.) of the glycoprotein mixture of interest (**Figs. 22 and 23**). The following six steps provides one example of how, a known or unknown glycopeptide mixture can be characterized using the techniques described herein.

Step 1:

- 74 -

Separate the glycans from the glycopeptide mixture. Isolate and sequence the resultant peptide(s). In this example, there was only one peptide chain and that was determined to be – YCNISQKMMSRNLTKDR. This peptide has two possible N glycosylation sites: CNIS and RNLT.

5 Step 2:

Digest the glycopeptide using trypsin followed by the cleavage of the glycans: one sample with ^{18}O labeling and another without labeling (^{16}O). Generate LC-MS spectra on both of the resultant samples. In this example, the following mass peaks were seen for the sample without labeling (289, 475, 476, 523, 855 and 856). With labeling the following mass peaks were seen (289, 475, 478, 523, 855, and 858).

10

By comparing the two spectra, the peptide fragments with mass 475 and 855 contain the glycosylation sites – both glycosylation sites are glycosylated. Based on a trypsin digest simulation of the peptide (YCNISQKMMSRNLTKDR) (See, for example, <http://us.expasy.org/tools/peptidecutter/>) the different masses were assigned as the following: 289 – DR; 475 – NLTK; 523 – MMSR; 855 – YCNISQK.

15

During the deglycosylation step, the Asn residue is converted into an Asp residue which results in a total increase in molecular weight of 1Da, thus explaining the appearance of the 476 and 856 peaks. The deglycosylation with concomitant ^{18}O -labeling, results in an increase of 2Da in the peptides that originally had a glycosylation site. This explains the appearance of the 478 and 858 peaks.

20

The quantitative measurement of the peaks via the methods described above reveals that the glycosylation site at NLTK is 75% glycosylated. Similarly, the data for YCNISQK reveals that it is 50% glycosylated. Similarly the undigested glycopeptide mixture is also cleaved of the glycans and label processed as described above. The resultant analysis shows that the entire mixture is 75% glycosylated.

25

Step 3:

The glycans are separated and the resultant glycans analyzed through MALDI-MS. In this example the resultant masses with Relative Abundance (**Table 8**) were:

30

Table 8 – Masses and Relative Abundance

| Mass | Relative Abundance |
|------|--------------------|
| 1235 | 40 |
| 1397 | 44 |
| 1559 | 16 |

Thus, there are three different glycans in this glycopeptide mixture.

Step 4:

- 75 -

Digest the glycopeptide mixture with trypsin and analyze the resultant mixture through MALDI-MS. In this example the resultant masses are 289, 475, 523, 854, 1871, 2033 and 2089.

Based on comparing the MS results with the trypsin digest simulation of the peptide, the following observations are made. Fragment NLTK is glycosylated with glycans with mass 1397 and 1559. Fragment YCNISQK is glycosylated with glycan with mass 1235.

Thus there are six possible glycopeptide chains in the mixture.

- Chain A that is not glycosylated.
- Chain B in which the second Asn is glycosylated with Glycan – 1397.
- Chain C in which the second Asn is glycosylated with Glycan – 1559.
- Chain D in which the first Asn is glycosylated with Glycan 1235.
- Chain E in which the first Asn is glycosylated with 1235 and the second with 1397.
- Chain F in which the first Asn is glycosylated with 1235 and the second with 1559.

Step 5:

Generate equations based on the experimental results and/or other data.

a,b,c,d,e and f are the relative abundances of the chains A, B, C, D, E and F respectively, and the following set of equations were generated based on the experimental results from steps 1 through 4.

| | | |
|----|-----------------------|---|
| | $a+b+c+d+e+f = 1$ | 6 possible chains |
| | $a+b+c = d+e+f$ | 50% occupancy in first glycosylation site |
| | $(a+d)*3 = (b+c+e+f)$ | 75% occupancy in second glycosylation site |
| 25 | $d+e+f = 2.5*(c+f)$ | Glycan 1235 to Glycan 1559 |
| | $b+e = 2.75*(c+f)$ | Glycan 1397 to Glycan 1559 |
| | $3*a = b+c+d+e+f$ | 75% of glycopeptide chains are glycosylated |

Solving the equations, the results are:

30 $a = .25, b = .25, c = d = 0, e = .3, f = .2$

Step 6:

The masses from step 3 can be resolved into potential glycan structures by using a glycan database lookup (<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/searchByMw.jsp>), and the exact structure of the carbohydrates were corroborated from the glycosidase digest analysis. By putting together the results in steps 1 to 6, the unknown glycoprotein mixture was determined to be (Tables 9 and 10):

Table 9 – Glycan Identification

| Peptide | Glycan Site 1 | Glycan Site 2 | Relative Abundance |
|---|---------------|---------------|--------------------|
| YCN ₁ ISQKMMSRN ₂ LTkDR | None | None | .25 |

| | | | |
|---|--------------------------------------|--------------------------------------|-----|
| YCN ₁ ISQKMMSRN ₂ LTKDR | None | HEX ₆ HEXNAC ₂ | .25 |
| YCN ₁ ISQKMMSRN ₂ LTKDR | HEX ₅ HEXNAC ₂ | HEX ₆ HEXNAC ₂ | .3 |
| YCN ₁ ISQKMMSRN ₂ LTKDR | HEX ₅ HEXNAC ₂ | HEX ₇ HEXNAC ₂ | .2 |

Table 10 – Glycan Structure

| Glycan | Structure |
|--------------------------------------|---|
| HEX ₅ HEXNAC ₂ | <pre> Man a1 \ 6 Man a1 / 3 Man a1 \ 6 / 3 Man b1—4 GlcNAc b1—4 GlcNAc \ 3 Man a1 / </pre> |
| HEX ₆ HEXNAC ₂ | <pre> Man a1 \ 6 Man a1 / 3 Man a1 \ 6 / 3 Man b1—4 GlcNAc b1—4 GlcNAc \ 3 Man a1—2 Man a1 / </pre> |
| HEX ₇ HEXNAC ₂ | <pre> Man a1—2 Man a1 \ 6 / 3 Man a1 \ 6 / 3 Man b1—4 GlcNAc b1—4 GlcNAc \ 3 Man a1—2 Man a1 / </pre> |

Analysis of glycosylation of glycoprotein standards

As an example, the optimized procedures were performed using two known *N*-glycosylated protein standards with different properties, ribonuclease B (RNaseB), a glycoprotein that only contains high mannose structures, and ovalbumin, which contains both hybrid and complex glycan structures at one glycosylation site. The procedures described above were applied to samples (obtained from the Hamel laboratory, MIT Bioprocess Engineering Center, Cambridge, MA) produced under various conditions.

Determination and quantification of glycosylation site occupancy

Before protease cleavage, the glycoproteins are first denatured in the presence of urea, reduced with DTT and carboxymethylated with iodoacetamide. To remove the denaturing reagents, the samples are concentrated using a centrifugal concentrator (3,000 MWCO) followed by buffer exchanged into protease compatible buffer (50 mM ammonium bicarbonate, pH 8.5, for trypsin digest). The proteins are then cleaved by proteases followed by denaturation of proteases by boiling the sample in water and lyophilization. Glycosylation site specific labeling is achieved by reacting the samples with PNGase F in the presence of

- 77 -

¹⁸O-water (**Fig. 24**). After desalting the glycosylated, unglycosylated, and ¹⁸O-labeled unglycosylated peptides through a C-18 solid phase extraction cartridge, these are used for LC-MS, LC-MS/MS, MALDI or MALDI-FTMS. For this study the ¹⁶O and ¹⁸O labeled samples were mixed in a 1:1 ratio before injection in order to facilitate the analysis. Other techniques for peptide sequencing can also be used at this point. The peptides were analyzed using a capillary LC-MS using a Vydac C-18 MS 5 µm (250x0.3mm) column coupled to a Mariner Biospectrometry Workstation. The peptides generated from the protease cleavage were corroborated using the Swiss-Prot database (ribonuclease B, P00656 and ovalbumin, P01012).

By studying the data obtained from the differentially labeled peptides after glycan cleavage, the specific glycosylation site can be easily determined. The introduction of the ¹⁸O at the glycosylation site is detected as a 2Da increase for a specific peptide. This data facilitate the determination of the glycosylation site and its occupancy. As determined using the peptide mass calculator from the protein data bank (<http://us.expasy.org/tools/peptide-mass.html>), the tryptic digest of ribonuclease B should yield a peptide fragment with a [M+H]⁺ of 475. 29 Da containing the glycosylation site (NLTK). Since the enzyme-mediated glycan cleavage generates an aspartic acid at the asparagine site, a peptide ion of [M+H]⁺ of 476. 29 Da for the unlabeled peptide and a 478.29 Da for the ¹⁸O-labeled peptide containing the glycosylation site was expected. As shown in **Fig. 25**, it was easy to identify the peptide fragment containing the glycosylation site in ribonuclease B by comparing the LC-MS data from the 1:1 mixture of ¹⁶O/¹⁸O labeled peptides against the unlabeled sample. The presence of the +2Da species in a 1:1 ratio in the ¹⁶O/¹⁸O labeled mixture and the absence of species with [M+H]⁺ of 475. 29 Da indicates that this peptide contains a glycosylation site and that it is 100% occupied in both samples of the mixture. By analyzing the differences between the peptide masses in this batch to the peptide masses from the samples not exposed to glycan cleavage, a preliminary identification of the glycans was obtained. This was further validated and quantified by analyzing the glycans separately as described below.

Release and purification of N-glycans from glycoprotein standards

Several enzymatic and chemical methods were used to separate glycans from their protein cores. Of the chemical methods, hydrazinolysis provides the most efficient release of glycans [31]. However, this approach requires the sample to be very clean, with no residual salts, and the reaction does not proceed efficiently in air or water, making hydrazinolysis

- 78 -

somewhat undesirable as a quick measure of quality control. N-glycanase F (PNGaseF) was chosen among enzymatic methods for the cleavage of N-linked glycans since the use of other enzymes results in loss of information such as fucosylation at the proximal GlcNAc.

For optimal enzyme activity proteins were unfolded, reduced and carboxymethylated prior to enzymatic digestion. Typically, the samples were denatured by heating in the presence of β -mercaptoethanol and/or SDS or by incubating at room temperature with urea, followed by reduction with DTT and carboxymethylation with iodoacetic acid or iodoacetamide.. To isolate the carbohydrates from the sample, the proteins were first precipitated with ethanol and the supernatant containing the glycans was then dried under vacuum and resuspended in water. Subsequent purification steps were required when detergents were used. Optimal results were obtained by using porous graphitic carbon columns. Neutral and charged carbohydrates were separated using these columns and eluted in mass spectrometry-compatible buffers. At this point, the most difficult component to get rid of was the detergent, which interferes with the types of analytical techniques that were used in this study.

Glycan Analysis

Different analytical techniques known in the art can be used for the glycan analysis methods. In this study MALDI-MS was used due to its simplicity and sensitivity (low femtomol after optimizations described above). The MALDI-MS protocol was optimized for the detection and quantification of low abundance carbohydrates (**Figs. 26 and 27**). In particular, **Fig. 27** shows the MALDI-MS spectrum of ovalbumin glycans. The observed peaks and their structures were found. The results are as shown above in **Table 3**.

RNAse B Computational analysis

The information obtained from the previous analysis was analyzed using the computational platform that contains the proteomics and glycomics based bioinformatics tools and databases described herein.

The sequence of the protein backbone was determined from the proteomics database as follows:

```
MALKSLVLLS LLVLVLLLVLR VQPSLGKETA AAKFERQHMD SSTSAASSSN YCNQMMKSRN1
LTKDRCKPVN TFFVHESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSKYPN
CAYKTTQANK HIIIVACEGNP YVPVHFDAVS
```

- 79 -

The glycosylation site is at SNLT. It is 100 % glycosylated and five different glycans were observed on analysis of the glycans via MALDI-MS. The results of the computational analysis showed that there were 5 different chains in the glycoprotein mixture as shown in **Table 11** below:

Table 11-Results from the Computational Analysis

| Protein Sequence | Glycan | Relative Abundance |
|---|--------------------------------------|--------------------|
| MALKSLVLLS LLVLVLLLVR VQPSLGKETA AAKFERQHMD SSTSAASSN YCNQMMKSRN ₁ LTKDRCKPVN TFFVESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSKYPN CAYKTTQANK HIIIVACEGNP YVPVHF ₂ DASV | HEX ₅ HEXNAC ₂ | .41 |
| MALKSLVLLS LLVLVLLLVR VQPSLGKETA AAKFERQHMD SSTSAASSN YCNQMMKSRN ₁ LTKDRCKPVN TFFVESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSKYPN CAYKTTQANK HIIIVACEGNP YVPVHF ₂ DASV | HEX ₆ HEXNAC ₂ | .29 |
| MALKSLVLLS LLVLVLLLVR VQPSLGKETA AAKFERQHMD SSTSAASSN YCNQMMKSRN ₁ LTKDRCKPVN TFFVESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSKYPN CAYKTTQANK HIIIVACEGNP YVPVHF ₂ DASV | HEX ₇ HEXNAC ₂ | .1 |
| MALKSLVLLS LLVLVLLLVR VQPSLGKETA AAKFERQHMD SSTSAASSN YCNQMMKSRN ₁ LTKDRCKPVN TFFVESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSKYPN CAYKTTQANK HIIIVACEGNP YVPVHF ₂ DASV | HEX ₈ HEXNAC ₂ | .14 |
| MALKSLVLLS LLVLVLLLVR VQPSLGKETA AAKFERQHMD SSTSAASSN YCNQMMKSRN ₁ LTKDRCKPVN TFFVESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSKYPN CAYKTTQANK HIIIVACEGNP YVPVHF ₂ DASV | HEX ₉ HEXNAC ₂ | .06 |

MALDI-MS Analysis Of N-Glycans from Antibodies Produced in Applikon and Wave Reactors

Two antibody samples produced by mouse-mouse hybridoma cells (Biokit SA, Barcelona, Spain) grown in an Applikon stirred tank reactor (STR) were analyzed, along with three samples produced in Wave reactors. The reactor conditions used are shown in **Table 12**.

Table 12. Reactor conditions used to produce antibody samples.

- 80 -

| Sample | Reactor Type | DO | pH | Other |
|--------|--------------|----------------|----------------|-----------------------------------|
| 1 | Applikon STR | 50% | 7 | |
| 2 | Applikon STR | 90% | Not controlled | |
| 3 | Wave | Controlled | Not controlled | |
| 4 | Wave | Controlled | 7 | NaHCO ₃ for pH control |
| 5 | Wave | Not controlled | 7 | Fresh media for pH control |

In the Applikon STR reactor, pH can be controlled automatically by the instrument, which dispenses CO₂, NaHCO₃ and O₂ as needed. In the Wave reactor, however, measurements must be taken manually and pH adjusted by hand. The pH in this reactor can be controlled by either adding fresh media as the cells grow, or adding NaHCO₃ for increased buffering capacity, and CO₂ as needed. The main difference between the reactor types is the mode of agitation. In the Applikon STR, a blade stirrer keeps the cell suspension in motion, while a sparger introduces oxygen to the system in a controlled manner. In the Wave reactor, a rocking motion generates waves that mix the components of the system and aids the transfer of oxygen and other gases into the system.

The purified antibodies were processed according to the optimized method described above. For each sample, 100µg of protein was used as the starting material. Both positive and negative ion modes were used in the MALDI-MS to determine whether there were charged sugars present. No acidic glycans were observed from the analysis; which indicated neutral sugars were obtained from the antibodies. The MALDI-MS data of the five antibody samples produced using different conditions contained the same six glycans with molecular weights of 1317 Da, 1463 Da, 1478 Da, 1625 Da, 1641 Da and 1787 Da. The corresponding structures to these glycans were determined using the methods described above and are shown in **Fig. 28** with their theoretical masses.

These results indicate that the production method did not alter the nature of the glycans present in the samples, rather, the quantities of some glycans were affected. Notably, samples prepared in the Wave reactor displayed a 40% decrease in 1625.4 Da glycan as well as 20% reductions in the 1787.7 Da glycan with respect to samples prepared in the Applikon reactor while the other glycans remained equal.

While the exact mechanisms for these changes are not known, it is interesting that the largest changes occurred due to reactor type, not reactor conditions such as pH, DO or media

- 81 -

composition. Because the Applikon STR and the Wave reactors differ most in their method of agitation, reactor configuration is therefore the most likely source of glycan variation.

Differences in protein glycosylation have been linked to shear stress, such as by the stirring blade or the gas sparger in an STR reactor. However, the turbulence created in the Wave reactor also generates shear stress. One hypothesis for the shear stress effect is that cells must increase their overall protein production in response to membrane and/or cytoskeletal damage. As a consequence, the biosynthetic enzymes for glycosylation are diverted away from the protein of interest [27].

Although most observed parameters, including total antibody production, were similar in Applikon STR and Wave cultures, cells from the Wave reactor had slight increases in metabolic rates. Changes in cell metabolism may yield effects similar to those caused by shear stress, as all glycoproteins synthesized in the cell must compete for the same machinery in the ER and golgi.

Example 3 References

1. Hirschberg, C.B., Snider, M.D. (1987) Topography of glycosylation in the rough endoplasmic reticulum and Golgi apparatus. *Annu Rev Biochem* **56**, 63-87.
2. Bause, E. (1983) Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem J* **209**, 331-6.
3. Marshall, R.D. (1972) Glycoproteins. *Annu Rev Biochem* **41**, 673-702.
4. Dwek, R.A. (1996) Glycobiology: Toward Understanding the Function of Sugars. *Chem Rev* **96**, 683-720.
5. O'Connor, S.E., Imperiali, B. (1996) Modulation of protein structure and function by asparagine-linked glycosylation. *Chem Biol* **3**, 803-12.
6. Crocker, P.R., Varki, A. (2001) Siglecs in the immune system. *Immunology* **103**, 137-45.
7. Helenius, A., Aebi, M. (2001) Intracellular functions of N-linked glycans. *Science* **291**, 2364-9.
8. Imperiali, B., O'Connor, S.E. (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol* **3**, 643-9.
9. Furukawa, K., Takamiya, K., Okada, M., Inoue, M., Fukumoto, S. (2001) Novel functions of complex carbohydrates elucidated by the mutant mice of glycosyltransferase genes. *Biochim Biophys Acta* **1525**, 1-12.

10. Jaeken, J., Matthijs, G. (2001) Congenital disorders of glycosylation. *Annu Rev Genomics Hum Genet* **2**, 129-51.
11. Freeze, H.H., Aebi, M. (1999) Molecular basis of carbohydrate-deficient glycoprotein syndromes type I with normal phosphomannomutase activity. *Biochim Biophys Acta* **1455**, 167-78.
12. Carchon, H., Van Schaftingen, E., Matthijs, G., Jaeken, J. (1999) Carbohydrate-deficient glycoprotein syndrome type IA (phosphomannomutase-deficiency). *Biochim Biophys Acta* **1455**, 155-65.
13. Powell, L.D., Sgroi, D., Sjoberg, E.R., Stamenkovic, I., Varki, A. (1993) Natural ligands of the B cell adhesion molecule CD22 beta carry N-linked oligosaccharides with alpha-2,6-linked sialic acids that are required for recognition. *J Biol Chem* **268**, 7019-27.
14. Sgroi, D., Varki, A., Braesch-Andersen, S., Stamenkovic, I. (1993) CD22, a B cell-specific immunoglobulin superfamily member, is a sialic acid-binding lectin. *J Biol Chem* **268**, 7011-8.
15. Karlsson, K.A. (1998) Meaning and therapeutic potential of microbial recognition of host glycoconjugates. *Mol Microbiol* **29**, 1-11.
16. Pritchett, T.J., Brossmer, R., Rose, U., Paulson, J.C. (1987) Recognition of monovalent sialosides by influenza virus H3 hemagglutinin. *Virology* **160**, 502-6.
17. Sears, P., Wong, C.H. (1998) Enzyme action in glycoprotein synthesis. *Cell Mol Life Sci* **54**, 223-52.
18. Varki, A. (1999) *Essentials of glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Parodi, A.J. (2000) Protein glucosylation and its role in protein folding. *Annu Rev Biochem* **69**, 69-93.
20. Chang, G.D., Chen, C.J., Lin, C.Y., Chen, H.C., Chen, H. (2003) Improvement of glycosylation in insect cells with mammalian glycosyltransferases. *J Biotechnol* **102**, 61-71.
21. Perlman, S., van den Hazel, B., Christiansen, J., Gram-Nielsen, S., Jeppesen, C.B., Andersen, K.V., Halkier, T., Okkels, S., Schambye, H.T. (2003) Glycosylation of an N-terminal extension prolongs the half-life and increases the in vivo activity of follicle stimulating hormone. *J Clin Endocrinol Metab* **88**, 3227-35.

22. Darling, R.J., Kuchibhotla, U., Glaesner, W., Micanovic, R., Witcher, D.R., Beals, J.M. (2002) Glycosylation of erythropoietin affects receptor binding kinetics: role of electrostatic interactions. *Biochemistry* **41**, 14524-31.
23. Krapp, S., Mimura, Y., Jefferis, R., Huber, R., Sondermann, P. (2003) Structural
5 analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J Mol Biol* **325**, 979-89.
24. Raju, T.S., Briggs, J.B., Borge, S.M., Jones, A.J. (2000) Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein
10 therapeutics. *Glycobiology* **10**, 477-86.
25. Kunkel, J.P., Jan, D.C., Butler, M., Jamieson, J.C. (2000) Comparisons of the glycosylation of a monoclonal antibody produced under nominally identical cell culture conditions in two different bioreactors. *Biotechnol Prog* **16**, 462-70.
26. Zhang, F., Saarinen, M.A., Itle, L.J., Lang, S.C., Murhammer, D.W., Linhardt, R.J.
15 (2002) The effect of dissolved oxygen (DO) concentration on the glycosylation of recombinant protein produced by the insect cell-baculovirus expression system. *Biotechnol Bioeng* **77**, 219-24.
27. Senger, R.S., Karim, M.N. (2003) Effect of Shear Stress on Intrinsic CHO Culture State and Glycosylation of Recombinant Tissue-Type Plasminogen Activator Protein.
20 *Biotechnol Prog* **19**, 1199-209.
28. Muthing, J., Kemminer, S.E., Conradt, H.S., Sagi, D., Nimtz, M., Karst, U., Peter-Katalinic, J. (2003) Effects of buffering conditions and culture pH on production rates and glycosylation of clinical phase I anti-melanoma mouse IgG3 monoclonal antibody R24. *Biotechnol Bioeng* **83**, 321-34.
29. Joao, H.C., Dwek, R.A. (1993) Effects of glycosylation on protein structure and
25 dynamics in ribonuclease B and some of its individual glycoforms. *Eur J Biochem* **218**, 239-44.
30. Harvey, D.J., Wing, D.R., Kuster, B., Wilson, I.B. (2000) Composition of N-linked carbohydrates from ovalbumin and co-purified glycoproteins. *J Am Soc Mass
30 Spectrom* **11**, 564-71.
31. Patel, T., Bruce, J., Merry, A., Bigge, C., Wormald, M., Jaques, A., Parekh, R. (1993) Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins. *Biochemistry* **32**, 679-93.

- 84 -

32. Tarentino, A.L., Plummer, T.H., Jr., Maley, F. (1974) The release of intact oligosaccharides from specific glycoproteins by endo-beta-N-acetylglucosaminidase H. *J Biol Chem* **249**, 818-24.
33. Tarentino, A.L., Maley, F. (1974) Purification and properties of an endo-beta-N-acetylglucosaminidase from *Streptomyces griseus*. *J Biol Chem* **249**, 811-7.
34. Tarentino, A.L., Gomez, C.M., Plummer, T.H., Jr. (1985) Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry* **24**, 4665-71.
35. Hu, G.F. (1995) Fluorophore-assisted carbohydrate electrophoresis technology and applications. *J Chromatogr A* **705**, 89-103.
36. Rhomberg, A.J., Ernst, S., Sasisekharan, R., Biemann, K. (1998) Mass spectrometric and capillary electrophoretic investigation of the enzymatic degradation of heparin-like glycosaminoglycans. *Proc Natl Acad Sci U S A* **95**, 4176-81.

Example 4-Glycome Profiling

Sample preparation and carbohydrate purification

Samples (usually 60 µl) from the different body fluids (serum, saliva, urine, tears, etc.) were processed in a similar manner as described below. Although in most cases the entire glycoproteome from the sample was analyzed, in some cases, the samples were further fractionated in order to analyze a “sub-glycome” from a specific body fluid. For example, a specific subset of proteins (such as antibodies, serum albumins, and other high abundance proteins) were removed from the original serum sample in order to analyze a more specific subset of glycoproteins in more detail. For fractionation, the sample proteome was divided into “high abundance” and “low abundance” using solid supports containing antibodies, proteins and synthetic molecules specific for the desired proteins to be removed or concentrated. For example, IgGs were removed using protein A agarose (Biorad), beads and serum albumin was removed using Affi-blue gel (Biorad). Other fractionations included the separation into acidic and basic proteome using cation and anion exchange chromatography or the separation between glycosylated and unglycosylated proteins using Con-A columns. The removal of specific proteins was quantified by western blots.

Proteins in the samples (either fractionated or unfractionated) were then denatured using a buffer containing 8M urea, 3.2 mM EDTA and 360 mM Tris, pH 8.6.[Papac, 1998]. Reduction and carboxymethylation of the sample proteome was then achieved using DTT and

iodoacetamide respectively. Although iodoacetic acid is often used as the alkylating agent for the carboxymethylation, this one is not optimal when analyzing body fluids containing a wide range of glycoproteins since it generally causes precipitation of most proteins. After removal of denaturing, reducing and alkylating reagents, *N*-glycans were selectively released from the glycoproteins by incubation with PNGase F. The steps for protein denaturing, protein alkylation and glycan release were also performed with the proteins bound to a solid support as described below. The released carbohydrates were then purified from the proteins and separated into neutrals and acidic glycans in one step using a graphitized carbon columns. The glycan purification step was also performed in a high-throughput format by using columns in 96-well plates. This process was facilitated by the use of a TECAN robot. This protocol allowed the processing of more than 90 samples at the same time.

Fractionation of Serum Proteins

As an example, to remove serum albumin and IgGs, Affi-Blue Gel (Biorad, 200 μ L) and Prot A (Biorad, 200 μ L) were mixed in a 1:1 ratio and placed in a serum protein column. The column was washed with 1mL of compatible serum protein binding buffer (20 mM phosphate, 100 mM NaCl, pH 7.2) using gravity flow. The column was placed in an empty 2mL collection tube and centrifuged at 10,000G for 20 seconds at 4°C. The flow was stopped during the sample preparation. Serum (60 μ L) was mixed with compatible serum protein binding buffer (180 μ L), and 200 μ L of diluted serum was added to the top of the resin bed and allowed to mix with the column for 15 minutes. The column was then centrifuged at 10,000G for 20 seconds at 4°C. Using the same collection tube, the column was washed with 200 μ L of compatible serum protein binding buffer and centrifuged again at 10,000G for 20 seconds at 4°C. For the removal of IgGs alone, only protein A agarose beads were used and the binding buffer was modified to 10 mM phosphate, 150 mM NaCl, pH 8.2.

To separate glycosylated (mainly high-mannose) from unglycosylated proteins, Concanavalin A-agarose beads (Vector Laboratories, Burlingame, CA) were used. To prepare the column, 3ml ConA-agarose slurry was washed with ConA buffer (20mM Tris, 1mM MgCl₂, 1mM CaCl₂, 500mM NaCl, pH 7.4). Before loading, 500 μ L of serum was mixed with 150 μ L of 5X ConA buffer. After washing with 3ml ConA buffer, glycoproteins were eluted with 2ml of 500mM α -methyl-mannoside and dialyzed against 10mM phosphate pH 7.2 overnight at 4°C.

- 86 -

Analysis of IgG and serum albumin depletion

Samples were prepared for SDS-Page electrophoresis by diluting 1:1 with 2X sample buffer (120 mM Tris base, 280 mM SDS, 20% Glycerol, 10% BME, 20 ng/ml BPB), boiled for 5 minutes, and 10ul was loaded per lane in a 4-12% Bis-Tris precast gel (Invitrogen: NPO323BOX). Lane one contained 5ul of a standard (BioRad: Precision All Blue Standard, 161-0373). The gel was run for 70 minutes at 200V. The gel was stained with SimplyBlue (Invitrogen:LC6060) according to the manufacturer. Imaging was performed on a Kodak Image Station 2000R. Another set of duplicate depleted samples were run as before and one gel was for SimplyBlue and the other was transferred to a 0.20um nitrocellulose membrane (Invitrogen:LC2000) employing an X Cell Blot Module (Invitrogen:E19051) for 70 minutes at 30V. The membrane was then blocked overnight at 4°C in 5% Blotto (Santa Cruz:sc-2325) and then probed with 1:1000 Protein A-hrp (Zymed :10-1023) for 1 hour at 4°C and washed 4 times with washing buffer (1xTBS:200mMTris base, 1.5M NaCl, pH7.5). The blot was developed with 4ml of substrate (ECL plus Western Blotting Detection System:RPN2132) for 2 minutes and then exposed. The bands corresponding to the treatments were manually captured as ROI (region of interest) employing the Kodak 1D Image Analysis Software and the Mean Intensity was normalized to the controls.

The same blot was then washed again and re-probed with 1:1000 Sheep Anti-human Albumin-hrp (Serotec:AHP102P) for 1 hour at 4°C. The blot was washed again, developed and imaged as before (**Fig. 29**).

Glycoblotting of Serum Samples

Protein samples were prepared for SDS-PAGE by diluting 1:1 with 2X denaturing buffer (40µg/ml SDS, 20% glycerol, 30µg/ml DTT and 10µg/ml bromophenol blue in 125mM Tris, pH 6.8) and boiling for 2min. Pre-cast Nu-PAGE 10% Bis-Tris protein gels were obtained from Invitrogen (Carlsbad, CA). Each lane was loaded with a maximum of 10µl of sample and run for 50min at 200V. After electrophoresis was complete, the gel was stained with Invitrogen SafeStain (1 hour in staining solution, then washed overnight with water).

The GlycoTrack glycoprotein detection kit was obtained from Prozyme (formerly Glyko). All reagents except buffers were supplied with the kit. Two methods were attempted—either biotinylating glycoproteins after blotting (a) or before blotting (b). For both methods, samples were first diluted 1:1 with 200mM sodium acetate buffer, pH 5.5.

- 87 -

The membrane was blocked by incubating overnight at 4°C with blocking reagent, then washed 3x10 minutes with TBS.

For method (a), samples were denatured with SDS sample buffer, and subjected to SDS-PAGE and blotting to nitrocellulose. After washing the membrane with PBS, the proteins were oxidized with 10ml of 10mM sodium periodate in the dark at room temperature for 20 minutes. The membrane was washed 3 times with PBS, and 2µl of biotin-hydrazide reagent was added in 10ml of 100mM sodium acetate, 2mg/ml EDTA for 60 minutes at room temperature. After 3 washes with TBS, the membrane was blocked overnight at 4°C with blocking reagent. Before adding 5µl of streptavidin-alkaline phosphatase (S-AP) conjugate, the membrane was washed again with TBS. The S-AP was allowed to incubate for 60 minutes at room temperature, and excess was washed off with TBS. To develop the blot, 50µl of nitro blue tetrazolium (50mg/ml) and 37.5µl of 5-bromo-4-chloro-3-indolyl phosphate p-toluidine (50mg/ml) were added in 10ml TBS, 10mg/ml MgCl₂. After 60 minutes, the blot was washed with distilled water and allowed to air dry.

In method (b), 20µl of sample was mixed with 10µl of 10mM periodate in 100mM sodium acetate, 2mg/ml EDTA and incubated in the dark at room temperature for 20 minutes. To destroy excess periodate 10µl of a 12.5mg/ml sodium bisulfite solution in 200mM NaOAc, pH 5.5 was added for 5 minutes at room temperature. Biotinylation was performed by adding 5µl of biotin amidocaproyl hydrazide solution in DMF. After incubating at room temperature for 60 minutes, the sample was mixed with SDS denaturing buffer and boiled for 2 minutes. Samples were run on SDS-PAGE gels and transferred to a nitrocellulose membrane (2 hrs, 30V). At this point, blocking and developing steps were identical to method (a) (**Fig. 30**).

Glycan release using solid supports: PNGaseF Digestion on PVDF Membrane

Glycans were also released using PVDF membranes as described by Papac, et.al. [Papac, 1998]. However, high abundance proteins were first removed before using this method since it resulted in low recoveries when processing entire body fluids. PVDF-coated wells in a 96-well plate were washed with 200µl MeOH, 3x 200µl H₂O and 200µl RCM buffer (8M urea, 360mM Tris, 3.2mM EDTA, pH 8.3). The protein samples (50µl) were then loaded in the wells along with 300µl RCM buffer. After washing with wells two times with fresh RCM buffer, 500µl of 0.1M DTT in RCM buffer was added for 1hr at 37°C. To

- 88 -

remove the excess DTT, the wells were washed three times with H₂O. For the carboxymethylation, 500µl of 0.1M iodoacetic acid in RCM buffer was added for 30 minutes at 37°C in the dark. The wells were washed again with water, then the membrane was blocked with 1ml polyvinylpyrrolidone (360,000 AMW, 1% solution in H₂O) for 1hr at room temperature. Before adding the PNGaseF, the wells were washed again with water. To release the glycans, 4µl of PNGaseF was added in 300µl of 50mM Tris, pH 7.5 and incubated overnight at 37°C. Released glycans were pipetted from the wells and purified through a graphitized carbon column. Similar to protocols used for the purification of glycans after performing the cleavage in solution, the purification of glycans after their release using PVDF membranes was also performed in a high-throughput format using columns in 96-well plates. This process was facilitated by the use of a TECAN robot.

Glycome analysis using mass spectrometry

Glycan analysis using methods known to the art were used and applied to the total body fluid glycome analysis. Using the methods provided above were also able to analyze more than 90 samples. Optimized MALDI-MS methods which did not required additional labeling and purification steps and also displayed great reproducibility and sensitivity for the carbohydrate analysis was used. As shown in **Fig. 31**, total serum glycome profiles typically displayed between 25-30 neutral glycans as well as 25-30 acidic glycans.

Using the look-up table described previously, almost all the peaks in MALDI-MS serum profiles could be identified as glycans of known composition. Many of the unidentified peaks are sodium adducts. The composition and mass of each labeled peak are as listed above in **Table 18**. However, a few peaks in the acidic glycan spectrum correspond to more than one composition. This is more common in the higher mass range since there are a larger number of possible monosaccharide compositions.

Validation of biomarker structures

MALDI-MS analysis allows us to analyze the entire glycome profile in a sample and compare the changes in the glycome composition between samples in a rapid and efficient manner. Due to the limitations in isomass characterization, in some instances, other techniques known to the art can be used to further characterize and validate the biomarkers determined from the total profile found using MALDI-MS techniques. For example, liquid chromatography-mass spectrometry (LC-MS) and capillary electrophoresis-laser induced

- 89 -

fluorescence (CE-LIF), are used in combination with a panel of exoglycosidases in order to obtain further linkage characterization of the carbohydrates (**Fig. 32**). After a specific pattern is established based on MALDI-MS results and the possible species are determined, matched samples displaying the differences in patterns are analyzed by these techniques in order to come up with defined structures of the biomarkers of interest. LC-MS/MS is also used to obtain linkage information based on the fragmentation patterns.

Other body fluids

Similar to the serum glycome analysis, the entire glycome from other body fluids such as saliva and urine have been studied (**Fig. 33**). For these cases, similar protocols employed for serum were used. In some instances, additional fractionations were also employed if a fraction of the glycome or glycoproteome was to be studied. The method showed to be equally, reproducible and sensitive for these other body fluids.

Glycome analysis of cell surface glycoproteins

The methods are also applied to the glycoprofiling of cell surfaces. All cell surface glycoproteins are cleaved using methods known to the art. Briefly, to harvest glycans using protease extraction, cells are washed 3X with PBS and incubated for 20-45 minutes with trypsin/EDTA (GibcoBRL) at 37°C for protease extraction. The samples are centrifuged for 10 minutes at 3000xg to pellet the cells, and the supernatant containing glycopeptides is collected and processed using methods described herein.

Glycomic pattern analysis

The emerging field of clinical proteomics has set new avenues for the identification of potential cancer-related biomarkers. In particular, the recent introduction of proteomic pattern diagnostics[Petricoin III, 2002;Wulfkühle, 2003; Conrads, 2003] provides a promising platform for the high-throughput discovery of new and important biomarkers. Since the alterations to the normal function of the glycosylation machinery have been increasingly recognized as a consistent indication of malignant transformations and tumorigenesis [Orntoft, 1999; Burchell, 2001; Brockhausen, 1999;Dennis, 1999] the final glycoproteins (specifically their carbohydrate moieties) can serve as sensitive and reliable biochemical markers to numerous diseases including cancer.

Some of the basic concepts from proteomic pattern diagnostics have been adapted and applied to total glycomic pattern analysis where the total profile of carbohydrates from body fluids or tissues can be examined in a rapid format. This approach provides an efficient overview of the total changes in carbohydrate composition of a tissue or body fluid as a result of pathological alterations and should be very reliable in sensing susceptible physiological changes to the body's natural homeostasis. This method not only serves as a fast diagnostic/prognostic tool but should also help to understand the function of specific carbohydrate modifications in some diseases. This method also provides a reliable system to efficiently monitor the effects of therapeutics.

The optimization of the MALDI-MS analysis allows reliable reproducibility that enables the fast evaluation of alterations to the glycomic patterns and their subsequent association to pathological/physiological changes to a sample donor. The optimized detection limits for this method (low femtomol) allows the detection of low abundance species associated to diseases. Every signal in the pattern is rapidly correlated to the glycan identity and can be further validated using a panel of glycosydases and other techniques. This prevents the erroneous identification as it has sometimes been the case in the field of proteomic pattern diagnostics. The pattern alterations can be easily determined manually or more efficiently with the aid of bioinformatics tools (described below). In some cases the decreasing levels of circulating glycoproteins in serum are easily matched to the analyzed glycans. As shown in **Fig. 34**, the glycome profile from serum with low IgG levels, reflects the specific decrease in the respective IgG glycans with molecular weights of 1463, 1626, 1666, 1788, 1829, 2102, and 1844. These glycans have been previously shown to be attached to IgG molecules in serum [Butler, 2003].

By applying the "glycomic pattern diagnostics" platform to different body fluids from patients with well-defined demographics, specific alterations in the glycomic pattern that can be correlated to the pathological state of the donor can be determined. For example, glycomic patterns have been associated to prostate cancer by studying the serum from prostate cancer patients (**Fig. 35**). Glycomic patterns from the saliva of patients with viral infections have also been established. Since every signal inside the pattern corresponds to specific glycans, the alteration of these patterns are easily determined and correlated with the expression levels of the carbohydrates, such as with the methods provided herein. The specific alterations in these glycan patterns are associated with a disease state. Therefore, the

methods provided serve as a reliable platform for diagnosis, prognosis and monitoring the effects of therapeutics.

Computational pattern analysis of glycoprofile

5 The following is an example of a computational approach for identifying glycan-based biomarkers for specific diseases using data from glycoprofilng. The different steps of the process are illustrated in the **Fig. 36**.

10 Using prostate cancer as an example, the goal is to identify glycan-based markers for individuals with prostate cancer. In this example, there are three possible categories of individuals – individuals with prostate cancer, benign prostatic hyperplasia (BPH) and individuals that are normal (healthy, non-diseased prostate).

15 Glycoprofilng data such as mass spectras are generated from samples from patients belonging to the different categories. Features are extracted from the glycoprofilng spectras. These features can be the presence or absence of one or more glycans in the profile, the relative amount of different glycans in the profile, combinations of different glycans found in the profile and other glycan-related properties. These glycans are identified in the glycoprofile spectra and can be corroborated with other methods, for instance, by using a Glycan database
20 (<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>) and/or associated glycomics-based bioinformatics tools.

25 The appropriate patient population is selected for the study (based on their history in a patient database), such that the subjects chosen in the different categories of prostate cancer, BPH and normal have the same distribution when it comes to other properties such as age, ethnicity, behavioral factors etc. This ensures that the variation in the glycan profiles can be attributed to the disease condition rather than other factors. The glycan related features extracted for this population via the previous step is run through a dataset generator to create the datasets needed for pattern analysis [see Weiss, S. & Indurkha, N. Predictive data mining - A practical guide, (Morgan Kaufmann, San Francisco, 1998)].

30 Different types of pattern analysis are performed to identify the patterns in this dataset [Weiss and Indurkha, 1998]. Three examples of patterns, rules or relationships that can be identified are as follows:

- 92 -

- *Linear Discriminant*: The pattern identified is in the form of weights (w_{11} , w_{12} , etc.) for the different glycan related features (G_1 , G_2 , etc.) as they are related to property or class of interest (ProstateCancer, BPH or Normal).

$$\begin{aligned} & \circ w_{11}G_1 + w_{12}G_2 + \dots + w_{1m}G_m + w_1 = \text{ProstateCancer} \\ & \circ w_{21}G_1 + w_{22}G_2 + \dots + w_{2m}G_m + w_2 = \text{BPH} \end{aligned}$$

- *Neural Network*: The neural network identifies non linear relationships or patterns between the different features and the property or class of interest.

$$\begin{aligned} & \circ \text{net}_j = \sum W_{ij} * f_i + C_j \\ & \circ d_j = 1 / (1 + e^{-\text{net}_j}), \text{ where } d_j \text{ can be Prostate Cancer, BPH or Normal} \end{aligned}$$

- *Decision Rules*: The pattern identified is in the form of *IF-THEN* rules, for example

(IF G_{11} is present and G_7 is not present) or (IF G_8 is present and G_9 is present) THEN Class = Prostate Cancer

(IF G_1 is present and G_2 is present and G_3 is not present) THEN Class = BPH

Otherwise Class = Normal

Once a pattern is identified using the decision set rules above, the patterns, rules or relationships are validated. The validation can be made based on variety of statistical methods that are used in biomarker validation as well as scientific methods to verify that the glycans found in the patterns do accurately reflect the disease state. If the patterns cannot be validated, the process described above can be repeated to look for other glycan-based patterns in the glycoprofiles.

Use of human glycome for the profiling of populations: Population-tailored treatments

It is well documented that different people react differently to certain drugs. In many instances this might be a result of drug interference with other inherent molecular components. Also, the down-regulation of enzymes may prevent the metabolism of some drugs or their by-products. The recent emphasis in the development of new carbohydrate-based therapeutics will face major challenges (in comparison to protein-based drugs) due to less currently available glycomic information as well as less understanding of the field.

More information of all human molecular components will significantly facilitate the design and prescription of medications to specific populations. The efficient analysis of the entire glycome from body fluids not only serves as a reliable diagnosis/prognosis platform but should become very valuable for the profiling of populations. The generation of a human glycome databank from different ethnic groups, gender, ages, diseases, etc. will become of enormous value for current and future development and applications of drugs that might interfere with carbohydrate functions. For example, the overexpression of specific

carbohydrates in a specific population will aid in the design or prescription of therapeutic antibodies (lectins, or other molecules) that might interfere with these molecules. This information will also aid in prospective studies for the selection of dosing, activity monitoring and efficacy endpoints.

5

References for Example 4

1. Petricoin III, E.F. et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359, 572-577 (2002).
2. Wulfkuhle, J.D., Liotta, L.A. & Petricoin III, E.F. Proteomic applications for the early
10 detection of cancer. *Nature Rev. Cancer* 3, 267-275 (2003).
3. Conrads, T.P., Zhou, M., Petricoin III, E.F., Liotta, L.A. & Veenstra, T.D. Cancer diagnosis using proteomic patterns. *Expert. Rev. Mol. Diagn.* 3, 411-420 (2003).
4. Orntoft, T.F. & Vestergaard, E.M. Clinical aspects of altered glycosylation of glycoproteins in cancer. *Electrophoresis* 20, 362-371 (1999).
- 15 5. Burchell, J.M., Mungul, A., . & Taylor-Papadimitriou, J. O-linked glycosylation in the mammary gland: changes that occur during malignancy. *J Mam. Gland Biol. Neoplasia* 6, 355-364 (2001).
6. Brockhausen, I. Pathways of O-glycan biosynthesis in cancer cells. *Biochim. Byophys. Acta* 1473, 67-95 (1999).
- 20 7. Dennis, J.W., Granovsky, M. & Waren, C.E. Glycoprotein glycosylation and cancer progression. *Biochim. Byophys. Acta* 1473, 21-34 (1999).
8. Pierce, M., Buckhaults, P., Chen, L. & Fregien, N. Regulation of N-acetylglucosaminyltransferase V and Asn-linked oligosaccharide beta(1,6) branching by a growth factor signaling pathway and effects on cell adhesion and metastatic
25 potential. *Glycoconj. J.* 14, 623-630 (1997).
9. Murata, K. et al. Expression of N-acetylglucosaminyltransferase V in colorectal cancer correlates with metastasis and poor prognosis. *Clin. Cancer Res.* 5, 1772-1777 (2000).
- 30 10. Yanagi, M., Aoyagi, Y., Suda, T., Mita, Y. & Asakura, H. N-Acetylglucosaminyltransferase V as a possible aid for the evaluation of tumor invasiveness in patients with hepatocellular carcinoma. *J. Gastroenterol. Hepatol.* 16, 1282-1289 (2001).

11. Ihara, S. et al. Prometastatic effect of N-acetylglucosaminyltransferase V is due to modification and stabilization of active matriptase by adding beta 1-6 GlcNAc branching. *J. Biol. Chem.* 277, 16960-16967 (2002).
12. Yousefi, S. et al. Increased UDP-GlcNAc:Gal beta 1-3GalNAc-R (GlcNAc to
5 GaLNac) beta-1, 6-N-acetylglucosaminyltransferase activity in metastatic murine tumor cell lines. Control of polylactosamine synthesis. *J. Biol. Chem.* 266, 1772-17782 (1991).
13. Shimodaira, K. et al. Carcinoma-associated expression of core 2 beta-1,6-N-acetylglucosaminyltransferase gene in human colorectal cancer: role of O-glycans in
10 tumor progression. *Cancer Res.* 57, 5201-5206 (1997).
14. Machida, E., Nakayama, J., Amano, J. & Fukuda, M. Clinicopathological significance of core 2 beta1,6-N-acetylglucosaminyltransferase messenger RNA expressed in the pulmonary adenocarcinoma determined by in situ hybridization. *Cancer Res.* 61, 2226-2231 (2001).
15. 15. Vander, A.J., Sherman, J.H. & Luciano, D.S. Human physiology : the mechanisms of body function, Edn. 8th. (McGraw-Hill, Boston, Mass.; 2001).
16. Rohovec, J., Maschmeyer, T., Aime, S. & Peters, J.A. The structure of the sugar residue in glycated human serum albumin and its molecular recognition by phenylboronate. *Chemistry* 9, 2193-2199 (2003).
- 20 17. Bihoreau, N., Ramon, C., Lazard, M. & Schmitter, J.M. Combination of capillary electrophoresis and matrix-assisted laser desorption ionization mass spectrometry for glycosylation analysis of a human monoclonal anti-Rhesus(D) antibody. *J Chromatogr B Biomed Sci Appl* 697, 123-133 (1997).
18. Watt, G.M. et al. Site-specific glycosylation of an aglycosylated human IgG1-Fc
25 antibody protein generates neoglycoproteins with enhanced function. *Chem Biol* 10, 807-814 (2003).
19. Papac, D.I., Briggs, J.B., Chin, E.T. & Jones, A.J. A high-throughput microscale method to release N-linked oligosaccharides from glycoproteins for matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis. *Glycobiology*
30 8, 445-454 (1998).
20. Butler, M. et al. Detailed glycan analysis of serum glycoproteins of patients with congenital disorders of glycosylation indicates the specific defective glycan

- 95 -

processing step and provides an insight into pathogenesis. Glycobiology 13, 601-622 (2003).

Each of the foregoing patents, patent applications and references that are recited in
5 this application are herein incorporated in their entirety by reference. Having described the
presently preferred embodiments, and in accordance with the present invention, it is believed
that other modifications, variations and changes will be suggested to those skilled in the art in
view of the teachings set forth herein. It is, therefore, to be understood that all such
variations, modifications, and changes are believed to fall within the scope of the present
10 invention as defined by the appended claims.

We claim:

- 96 -

CLAIMS

1. A method of analyzing a sample containing glycoconjugates, comprising:

(a) separating the glycans from the sample containing the glycoconjugates,

(b) determining the glycosylation sites and glycosylation site occupancy of the glycoconjugates,

(c) analyzing the glycans to characterize the glycans, and

(d) determining the glycoforms of the glycoconjugates in the sample with the results obtained from steps (b) and (c) with a computational method.

2. The method of claim 1, wherein step (a) comprises denaturing the glycoconjugates.

3. The method of claim 2, wherein the glycoconjugates are denatured with a denaturing agent.

4. The method of claim 3, wherein the denaturing agent is detergent, urea, guanidium hydrochloride or heat.

5. The method of claim 3, wherein the glycoconjugates are reduced following their denaturation.

6. The method of claim 5, wherein the glycoconjugates are reduced with a reducing agent.

7. The method of claim 6, wherein the reducing agent is DTT, β -mercaptoethanol or TCEP.

8. The method of claim 5, wherein the glycoconjugates are alkylated with an alkylating agent following their reduction.

9. The method of claim 8, wherein the alkylating agent is iodoacetic acid or iodoacetamide.

10. The method of claim 1, wherein the step of determining the glycosylation sites and glycosylation site occupancy comprises analyzing the glycoconjugates with 2D-NMR.

- 97 -

11. The method of claim 1, wherein the glycoconjugates comprise a peptide and wherein the step of determining the glycosylation sites and glycosylation site occupancy, comprises :

cleaving the backbone of the glycoconjugates,

cleaving and labeling with a first label the glycoconjugates of a first portion of the

5 sample at their glycosylation sites,

cleaving the glycoconjugates of a second portion of the sample at their glycosylation sites,

analyzing the first and second portions of the sample of glycoconjugates, and

quantifying the results of the analysis step.

10

12. The method of claim 11, wherein the glycoconjugates of the first portion are labeled with a labeling agent.

13. The method of claim 12, wherein the labeling agent is an isotope of C, N, H, S or O.

15

14. The method of claim 13, wherein the label is O^{18} .

15. The method of claim 11, wherein the glycoconjugates of the second portion are unlabeled.

20

16. The method of claim 11, wherein the glycoconjugates of the second portion are labeled.

17. The method of claim 11, wherein the first and second portions of the sample of glycoconjugates are analyzed separately.

25

18. The method of claim 11, wherein the first and second portions of the sample of glycoconjugates are analyzed as a mixture.

19. The method of claim 11, wherein the first and second portions of the sample are analyzed with a mass spectrometric method.

30

20. The method of claim 11, wherein the glycosylation site occupancy is quantified from ratios of the masses of cleaved glycoconjugates of the first and second portions of the sample.

- 98 -

21. The method of claim 19, wherein the mass spectrometric method is LC-MS, LC-MS/-MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.

5 22. The method of claim 11, wherein the step further comprises generating a list of possible glycoconjugates.

23. The method of claim 11, wherein the step further comprises generating a list of possible glycans.

10

24. The method of claim 1, wherein the step of analyzing the glycans comprises analyzing the glycans with a mass spectrometric method, an electrophoretic method, NMR, a chromatographic method or a combination thereof.

15 25. The method of claim 24, wherein the mass spectrometric method is LC-MS, LC-MS/MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.

26. The method of claim 25, wherein the MALDI-MS is MALDI-MS optimized with a mixture of 6-aza-2-thiothymine (ATT) and Nafion coating.

20

27. The method of claim 24, wherein the electrophoretic method is CE-LIF.

28. The method of claim 24, wherein the step further comprises contacting the glycans with one or more glycan-degrading enzymes.

25

29. The method of claim 28, wherein the one or more glycan-degrading enzymes is sialidase, galactosidase, mannosidase, N-acetylglucosaminidase or a combination thereof.

30 30. The method of claim 1, wherein the step of analyzing the glycans comprises quantifying the glycans using calibration curves of known glycan standards.

31. The method of claim 1, wherein the glycoconjugate comprises a peptide and wherein the method further comprises determining a peptide sequence of the glycoconjugate.

- 99 -

32. A method of analyzing a sample containing glycoconjugates, comprising:
separating the glycans from the sample containing the glycoconjugates,
determining the glycosylation sites and glycosylation site occupancy of the
glycoconjugates, and
analyzing the glycans to characterize the glycans,
wherein determining the glycosylation sites and glycosylation site occupancy
comprises cleaving and labeling with a first label the glycoconjugates at their glycosylation
sites of a first portion of the sample, cleaving the glycoconjugates at their glycosylation sites
of a second portion of the sample, analyzing the first and second portions of the sample of
glycoconjugates, and quantifying the results.

33. The method of claim 32, wherein the glycoconjugates of the first portion are labeled.

34. The method of claim 32, wherein the glycoconjugates of the second portion are
unlabeled.

35. The method of claim 32, wherein the glycoconjugates of the second portion are labeled.

36. The method of claim 32, wherein the first and second portions of the sample of
glycoconjugates are analyzed with a mass spectrometric method.

37. The method of claim 36, wherein the mass spectrometric method is LC-MS, LC-MS/MS,
MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.

38. The method of claim 32, wherein determining the glycosylation sites and glycosylation
site occupancy further comprises generating a list of possible glycoconjugates.

39. The method of claim 32, wherein the step of separating the glycans from the sample
comprises denaturing the glycoconjugates with a denaturing agent.

40. The method of claim 39, wherein the glycoconjugates are reduced with a reducing agent
following their denaturation.

- 100 -

41. The method of claim 40, wherein the glycoconjugates are alkylated with an alkylating agent following their reduction.

5 42. The method of claim 32, wherein the step of analyzing the glycans comprises analyzing the glycans with a mass spectrometric method, an electrophoretic method, NMR, a chromatographic method or some combination thereof.

43. The method of claim 42, where the mass spectrometric method is LC-MS, LC-MS/MS,
10 MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.

44. The method of claim 42, wherein the electrophoretic method is CE-LIF.

45. The method of claim 42, wherein the step further comprises contacting the glycans with
15 one or more glycan-degrading enzymes.

46. The method of claim 32, wherein the glycoconjugate comprises a peptide and wherein the method further comprises determining a peptide sequence of the glycoconjugate.

20 47. A method of determining the glycosylation site occupancy of glycoconjugates in a sample, comprising:

cleaving and labeling with a first label the glycoconjugates at their glycosylation sites of a first portion of the sample, cleaving the glycoconjugates at their glycosylation sites of a second portion of the sample, analyzing the first and second portions of the sample of
25 glycoconjugates, and quantifying the results.

48. The method of claim 47, wherein the method further comprises determining the possible fragments of the glycoconjugate.

30 49. The method of claim 47, wherein the glycoconjugates of the first portion are labeled with an isotope of C, N, H, S or O.

50. The method of claim 49, wherein the label is O¹⁸.

51. The method of claim 47, wherein the glycoconjugates of the second portion are unlabeled.

5 52. The method of claim 47, wherein the glycoconjugates of the second portion are labeled.

53. The method of claim 47, wherein the first and second portions of the sample of glycoconjugates are analyzed with a mass spectrometric method.

10 54. The method of claim 53, wherein the mass spectrometric method is LC-MS, LC-MS/MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.

55. A method of analyzing a sample containing glycans, comprising:

separating neutral from charged glycans, and

15 analyzing the neutral and charged glycans separately to characterize the glycan.

56. The method of claim 55, wherein the analyzing comprises analyzing the glycans with MALDI-MS.

20 57. A method of analyzing a glycan, comprising:

analyzing the glycan in the presence of Nafion and ATT.

58. The method of any one of claims 1, 32, 47, 55 and 57, wherein the method is a method of analyzing the purity of a sample containing glycans.

25 59. The method of any one of claims 1, 32, 47, 55 and 57, wherein the method is a method of analyzing the glycans of a sample of one or more cells, a tissue or body fluid from a subject.

60. A high-throughput method of any one of claims 1, 32, 47, 55 and 57, wherein more than
30 one sample is analyzed.

61. The method of claim 60, wherein the more than one sample of glycoconjugates are contained in a 96-well plate.

- 102 -

62. The method of claim 60, wherein the more than one sample of glycoconjugates are affixed to a membrane.

5 63. A method of generating a glycopeptide library, comprising:

cleaving the backbone of the glycopeptides in a sample and labeling the glycopeptide fragments generated with a labeling agent, and

cleaving the glycans in the sample.

10 64. The method of claim 63, wherein the labeling agent is O¹⁸.

65. The method of claim 64, wherein the labeling agent is an isotope of C, N, H, S or O.

15 66. The method of claim 63, wherein the method further comprises characterizing the fragments generated from the cleavage of the glycopeptides.

67. The method of claim 66, wherein the characterization is performed with LC-MS, LC-MS/MS, MALDI-MS, MALDI-TOF, TANDEM-MS or FTMS.

20 68. The method of claim 66, wherein the characterizing comprises characterizing the glycosylation sites, characterizing the peptides of the glycopeptides and characterizing the glycans.

69. A library of glycopeptides generated with the method of claim 63.

25

70. A method of analyzing a sample of glycopeptides, comprising:

characterizing the glycopeptides in the sample, and

comparing the characterized glycopeptides with the library of glycopeptides of claim

69.

30

71. A method of generating a list of glycoconjugate properties, comprising:

measuring two or more properties of the glycoconjugate, and

recording a value for the two or more properties of the glycoconjugate to generate a list, wherein the value of the two or more properties is recorded in a computer-generated data structure.

5 72. The method of claim 71, wherein one of the two or more properties of the glycoconjugates is the number of one or more types of monosaccharides of the glycoconjugate.

73. The method of claim 71, wherein one of the two or more properties of the
10 glycoconjugates is the total mass of the glycans of the glycoconjugate.

74. The method of claim 71, wherein the glycoconjugate is a glycoprotein or proteoglycan, and wherein the one of the two or more properties of the glycoconjugate is the mass of the peptide of the glycoconjugate.

15 75. The method of claim 71, wherein the glycoconjugate is a glycolipid, and wherein one of the two or more properties of the glycoconjugate is the mass of the lipid of the glycoconjugate.

20 76. The method of claim 71, wherein one of the two or more properties of the glycoconjugate is the mass of the glycoconjugate.

77. The method of claim 71, wherein one of the two or more properties of the glycoconjugate is the mass of permethylated glycans.

25 78. A database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more glycoconjugates, the database comprising:

one or more data units corresponding to the one or more glycoconjugates, each of the data units including an identifier that includes two or more fields, each field for storing a
30 value corresponding to one or more properties of the glycoconjugates.

79. A method of analyzing the total glycome of a sample of body fluid, comprising:
(a) analyzing the glycans of the sample, and

- 104 -

(b) determining a profile of the glycans of the sample.

80. The method of claim 79, wherein the method further comprises performing a pattern analysis on the results from (a) using a computational method.

5

81. The method of claim 79, wherein step (a) includes quantifying the glycans using calibration curves of known glycan standards.

82. The method of claim 80, wherein the method further includes recording the pattern in a computer-generated data structure.

10

83. The method of claim 79, wherein the profile of the glycans provides diagnostic or prognostic information.

84. The method of claim 79, wherein the method is a method for assessing the purity of the sample.

15

85. The method of claim 79, wherein the sample is a sample of serum, plasma, blood, urine, saliva, sputum, tears, CSF, seminal fluid, feces, tissues or cells.

20

86. A method of analyzing the total glycome of a sample of body fluid, comprising:

(a) analyzing the glycans of a sample of body fluid, and

(b) comparing the results from (a) with a known pattern.

87. The method of claim 86, wherein the sample is a sample of serum, plasma, blood, urine, saliva, sputum, tears, CSF, seminal fluid, feces, tissues or cells.

25

88. The method of claim 86, wherein the method is a method of diagnosis and the pattern is associated with a diseased state.

30

89. The method of claim 88, wherein the pattern associated with a diseased state is a pattern associated with cancer.

- 105 -

90. The method of claim 89, wherein the cancer is prostate cancer, melanoma, bladder cancer, breast cancer, lymphoma, ovarian cancer, lung cancer, colorectal cancer or head and neck cancer.

5 91. The method of claim 88, wherein pattern associated with a diseased state is a pattern associated with an immunological disorder.

92. The method of claim 88, wherein pattern associated with a diseased state is a pattern associated with a neurodegenerative disease.

10

93. The method of claim 92, wherein the neurodegenerative disease is a transmissible spongiform encephalopathy, Alzheimer's disease or neuropathy.

15 94. The method of claim 88, wherein pattern associated with a diseased state is a pattern associated with inflammation.

95. The method of claim 88, wherein pattern associated with a diseased state is a pattern associated with rheumatoid arthritis.

20 96. The method of claim 88, wherein pattern associated with a diseased state is a pattern associated with cystic fibrosis.

97. The method of claim 88, wherein pattern associated with a diseased state is a pattern associated with an infection.

25

98. The method of claim 97, wherein the infection is viral or bacterial.

99. The method of claim 86, wherein the method is a method of monitoring prognosis and the known pattern is associated with the prognosis of a disease.

30

100. The method of claim 86, wherein the method is a method of monitoring drug treatment and the known pattern is associated with the drug treatment.

- 106 -

101. A method of determining the purity of a sample, comprising:

- (a) analyzing the total glycans of the sample,
- (b) identifying the glycan pattern of the sample, and
- (b) comparing the pattern with a known pattern of a sample of predetermined purity to

5 assess the purity of the sample.

102. A method of generating the complete glycan profile of a body fluid, comprising:

- (a) analyzing the glycans in the sample, and
- (b) identifying the complete glycan profile of the sample.

10

103. The method of claim 102, wherein neutral, charged, N-linked and O-linked glycans are included in the profile.

104. The method of claim 102, wherein the sample is a sample of serum, plasma, blood,
15 urine, saliva, sputum, tears, CSF, seminal fluid, feces, tissues or cells.

105. A method of analyzing the total glycome of a sample, comprising:

determining the glycosylation site and glycosylation site occupancy of all
glycoconjugates in the sample,

20 characterizing components of the glycoconjugates and all glycans of the glycome in
the sample,

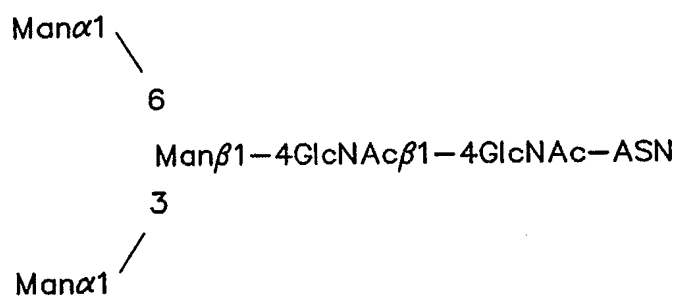
and matching specific glycans to glycoconjugates with a computational method.

106. A method of analyzing a sample of glycoconjugates, comprising:

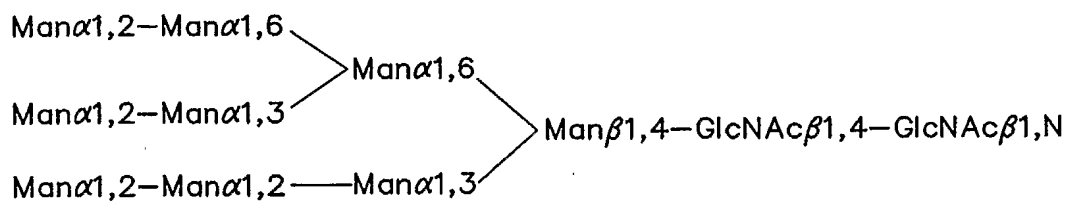
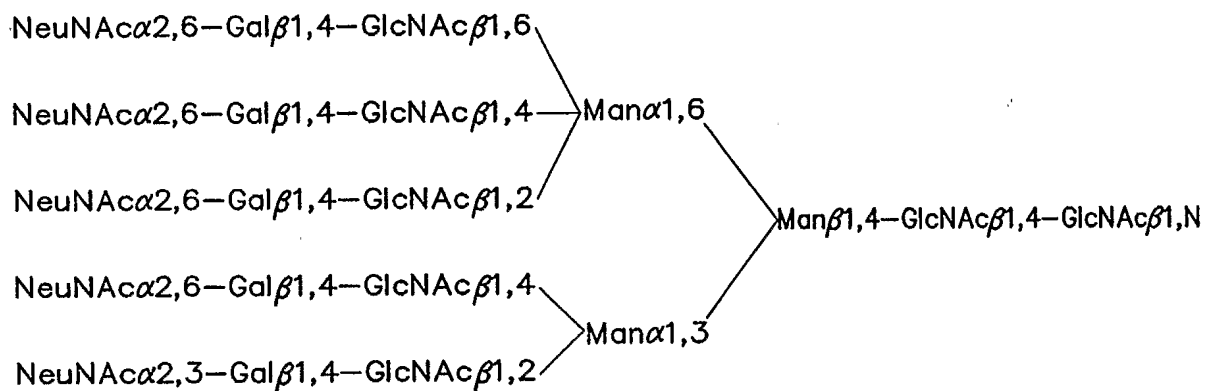
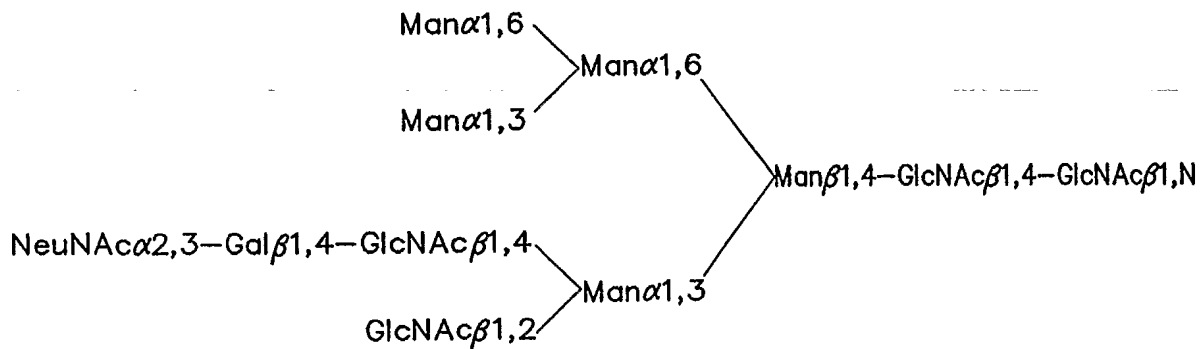
25 analyzing the glycans of the sample with an analytical method, and
determining the glycoforms of the sample with a computational method.

107. The method of claim 1 or 106, wherein the computational method comprises generating constraints from the experimental analysis and solving them.

1/60

**FIG. 1**

2/60

*High Mannose Type***FIG. 2A***Complex Type***FIG. 2B***Hybrid Type***FIG. 2C**

N-Glycans Biosynthesis

3/60

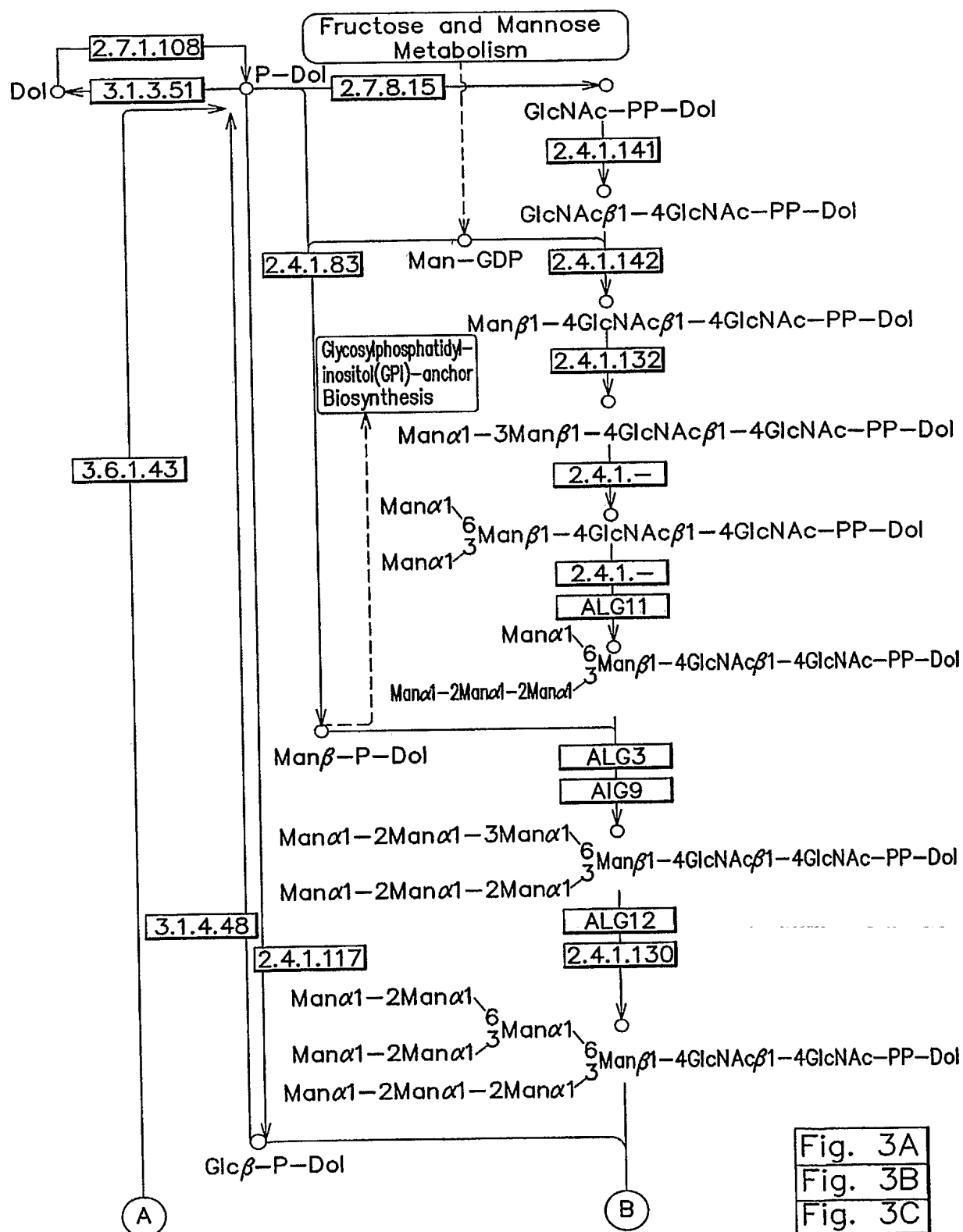


Fig. 3A
 Fig. 3B
 Fig. 3C
 Fig. 3D

FIG. 3A

4/60

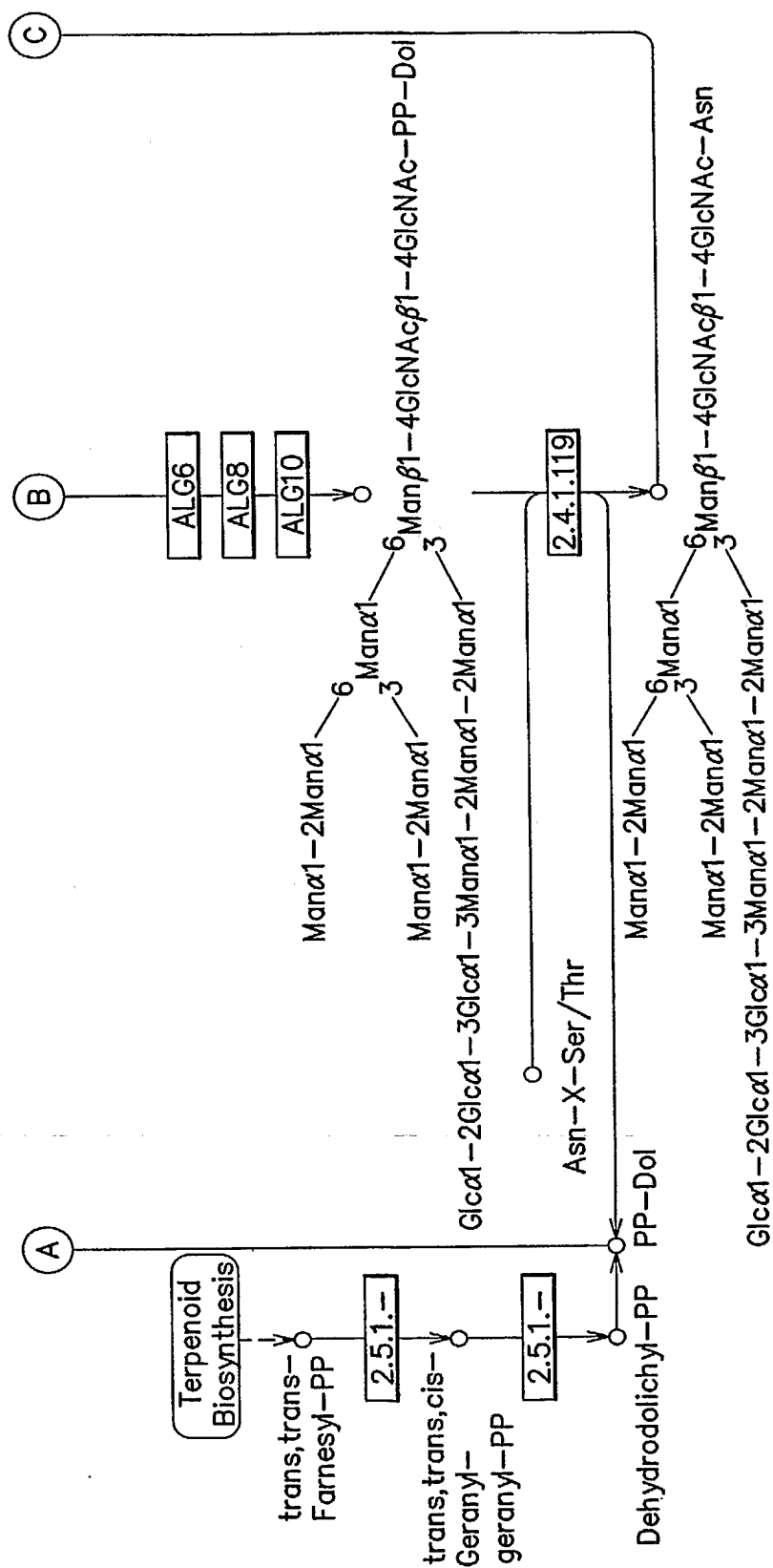


FIG. 3B

5/60

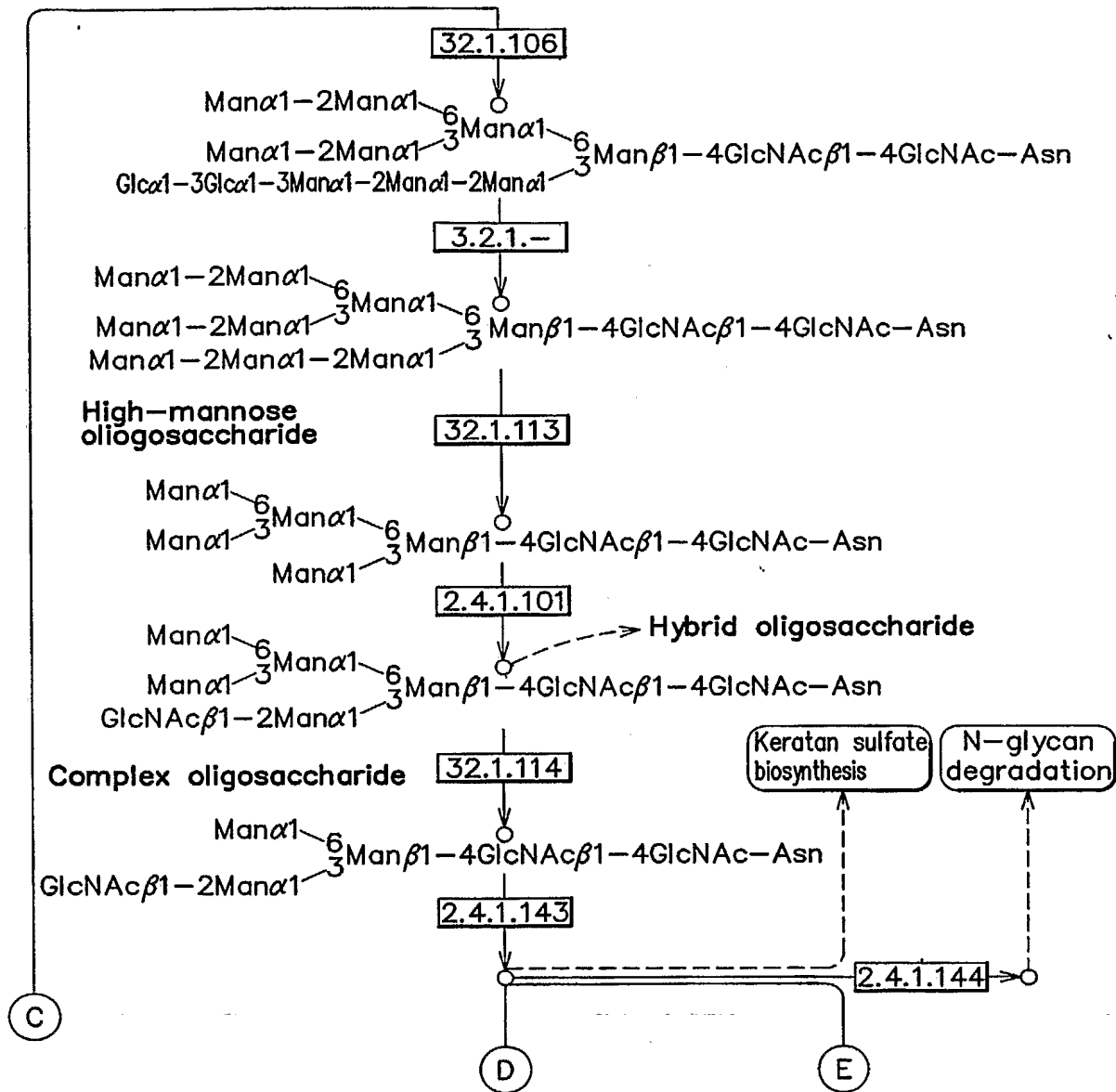


FIG. 3C

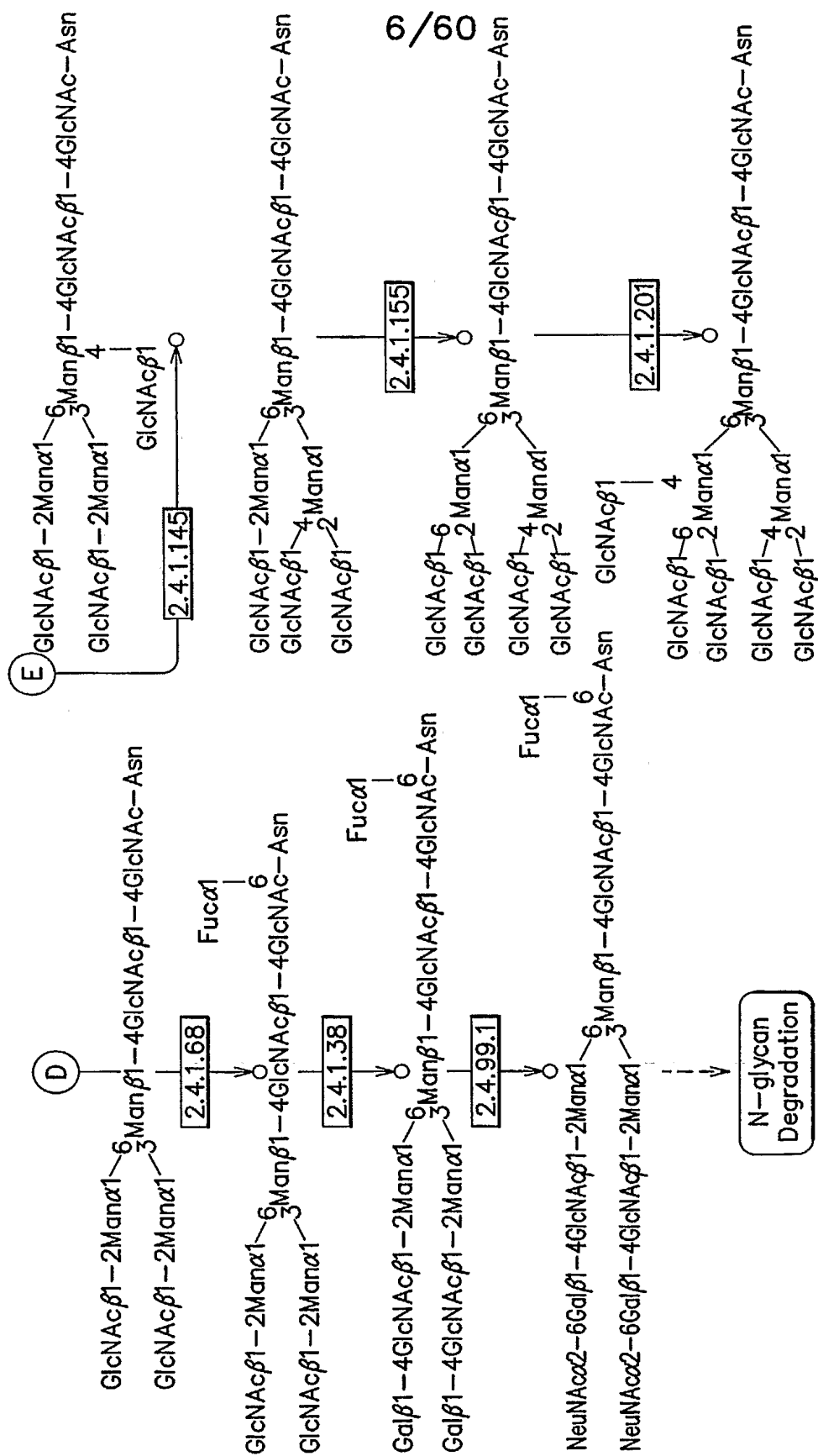
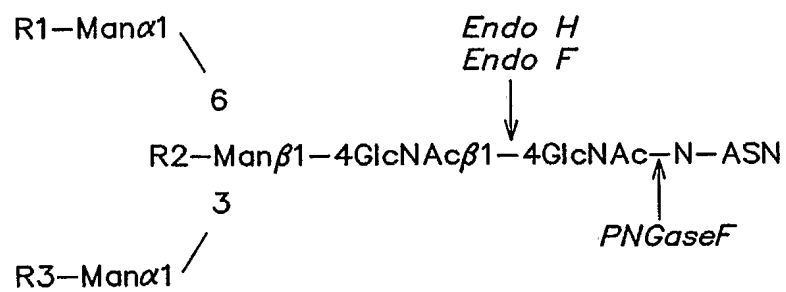


FIG. 3D

7/60

**FIG. 4**

8/60

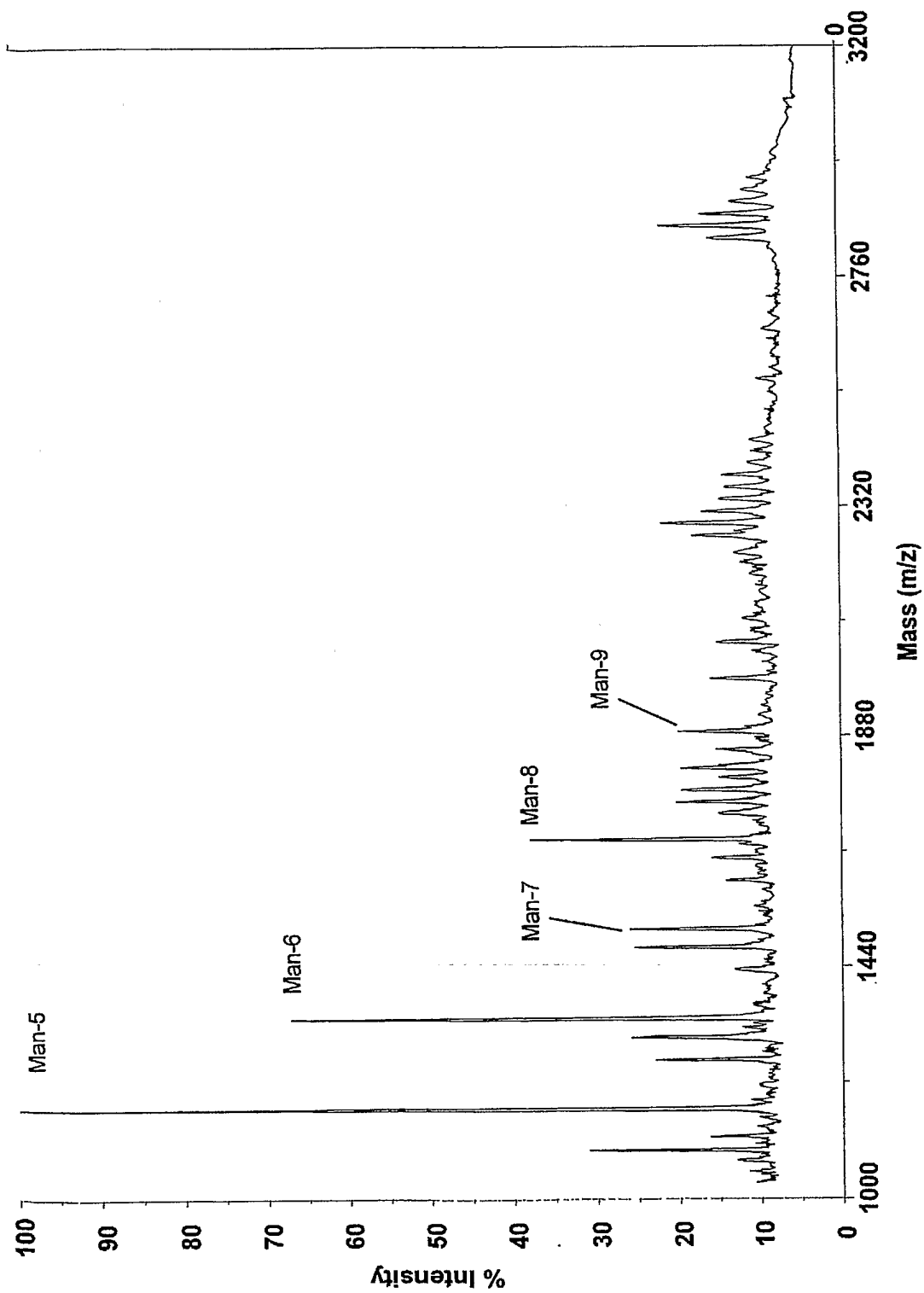


FIG. 5A

9/60

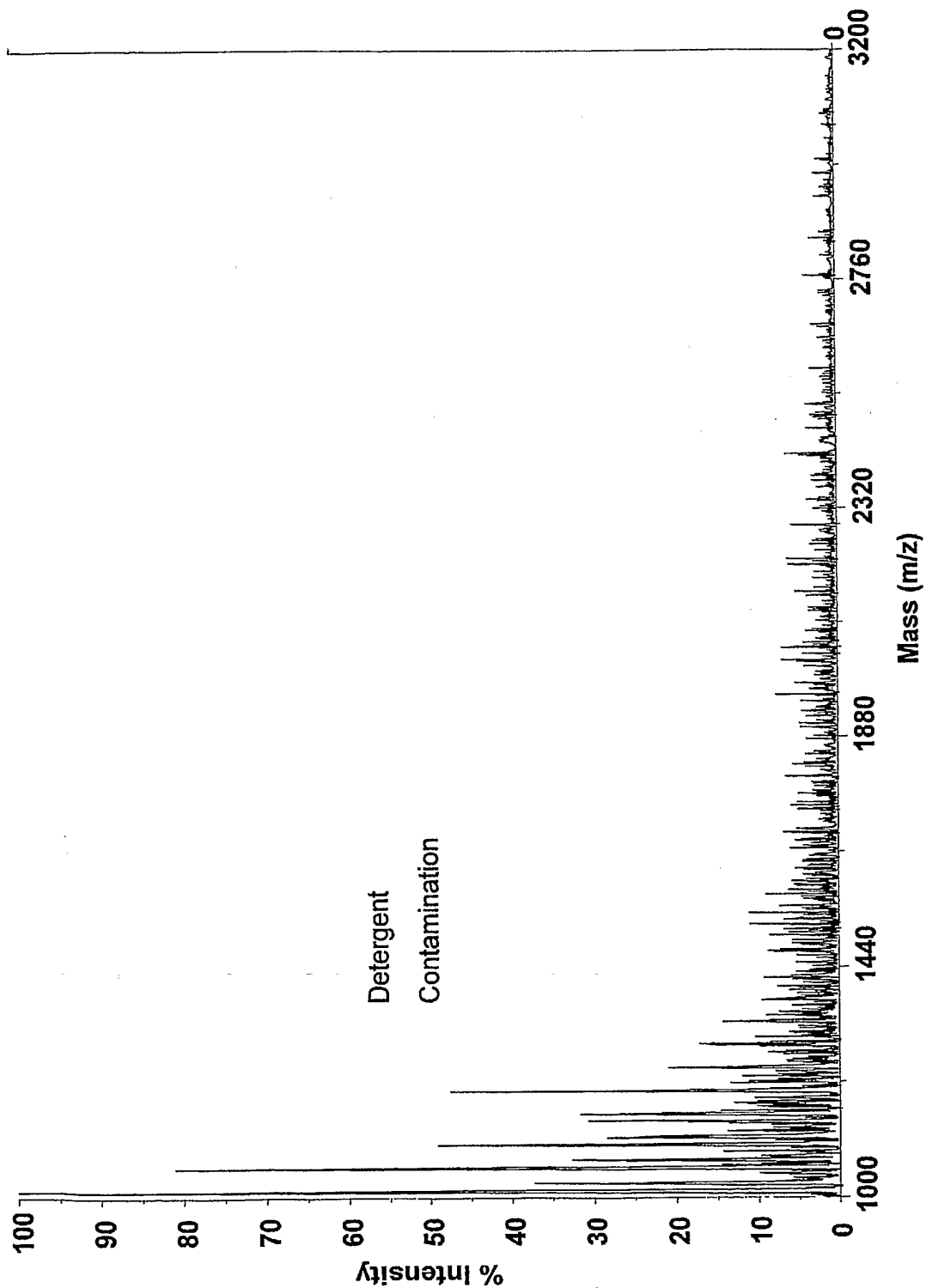
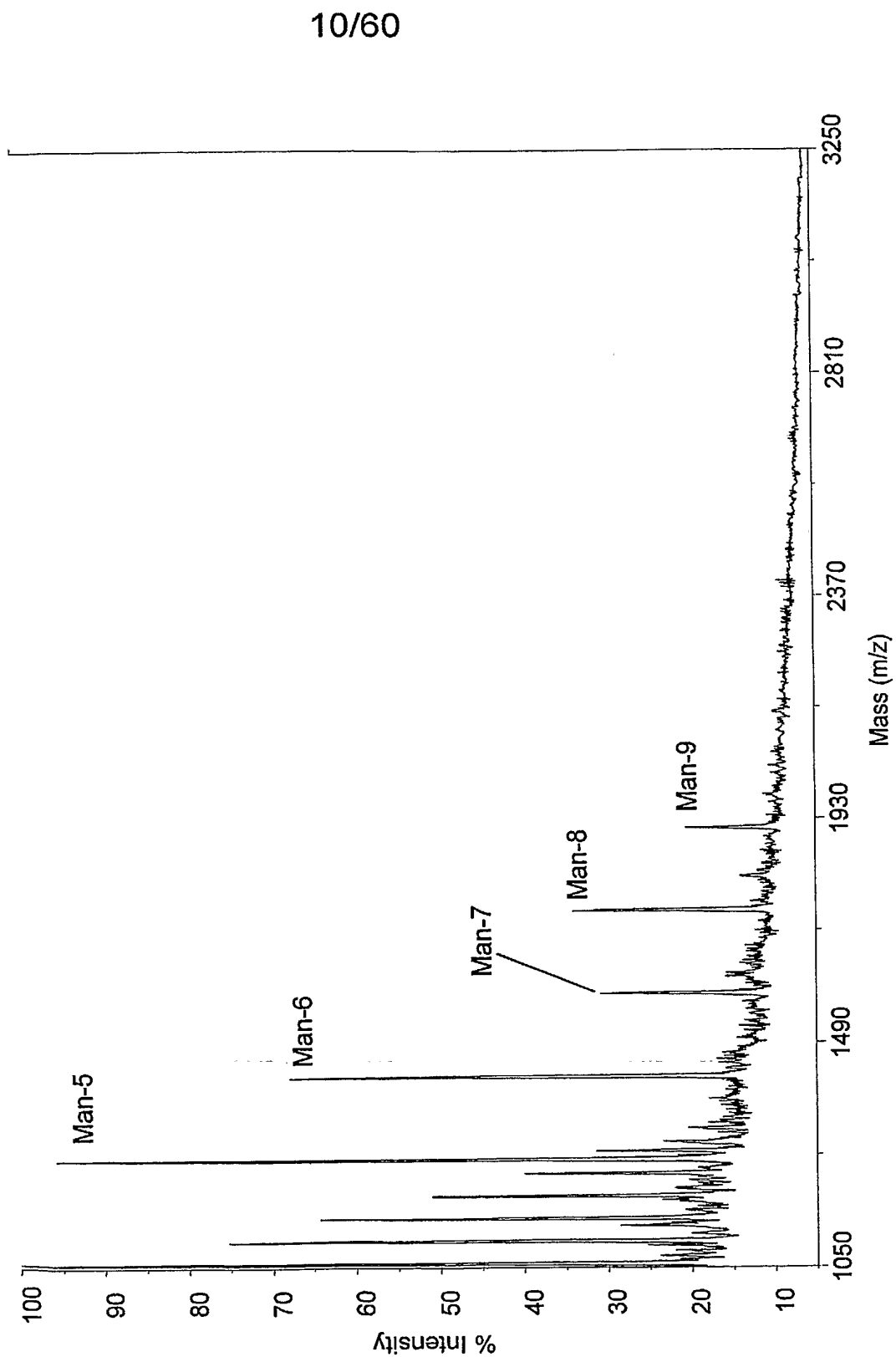


FIG. 5B

**FIG. 5C**

11/60

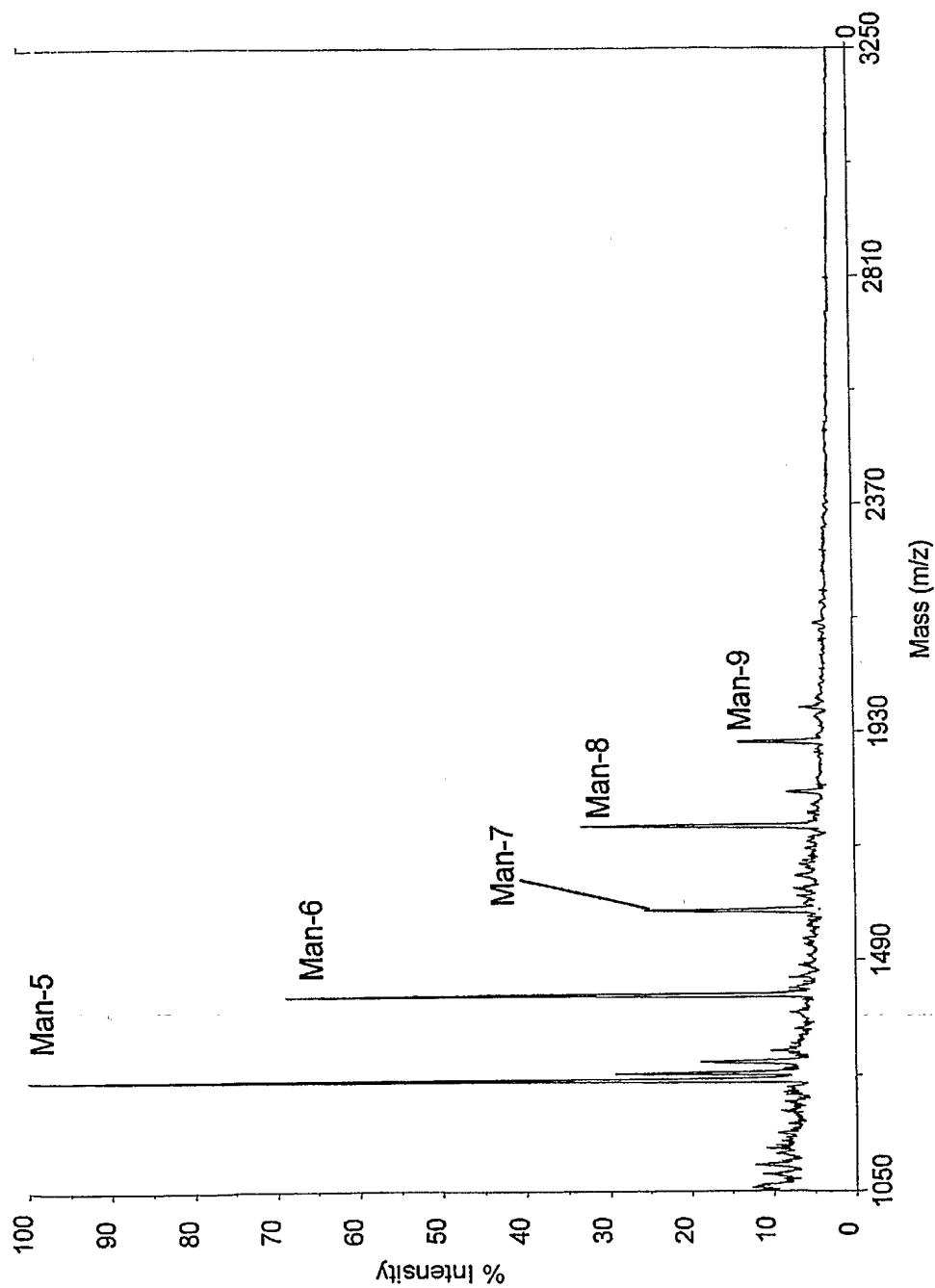
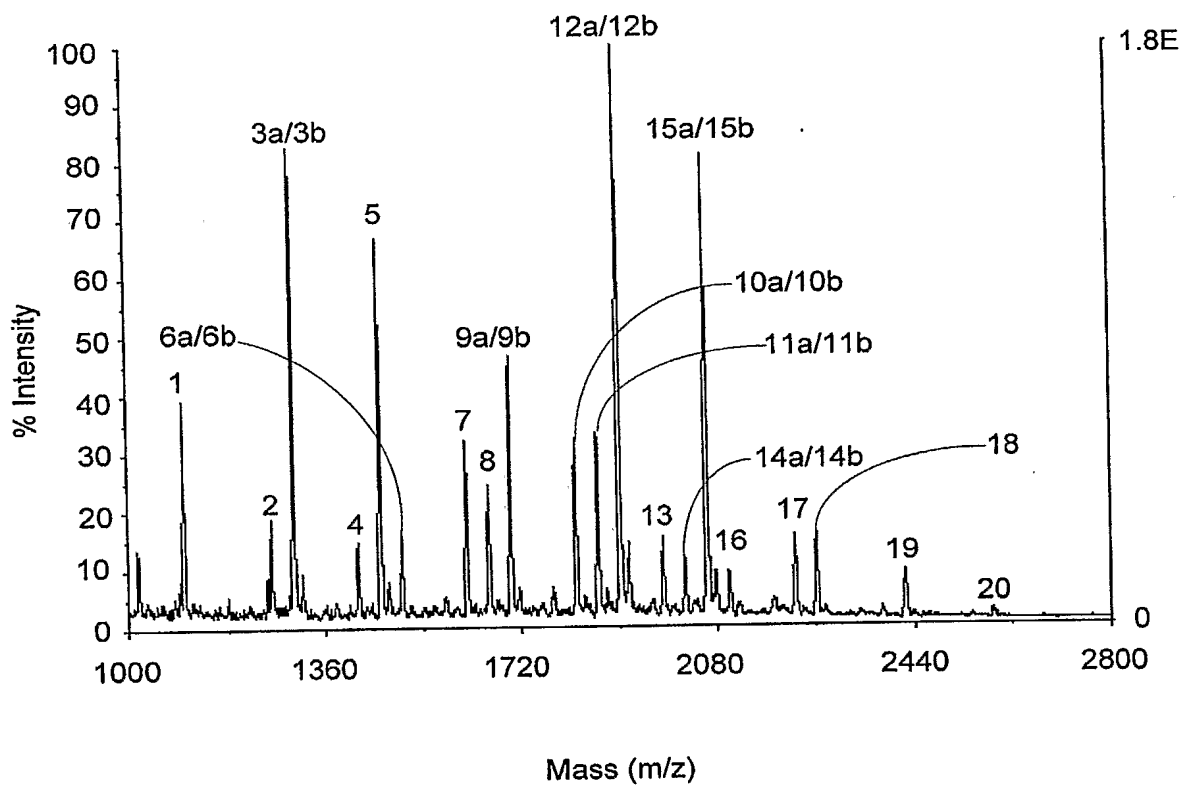
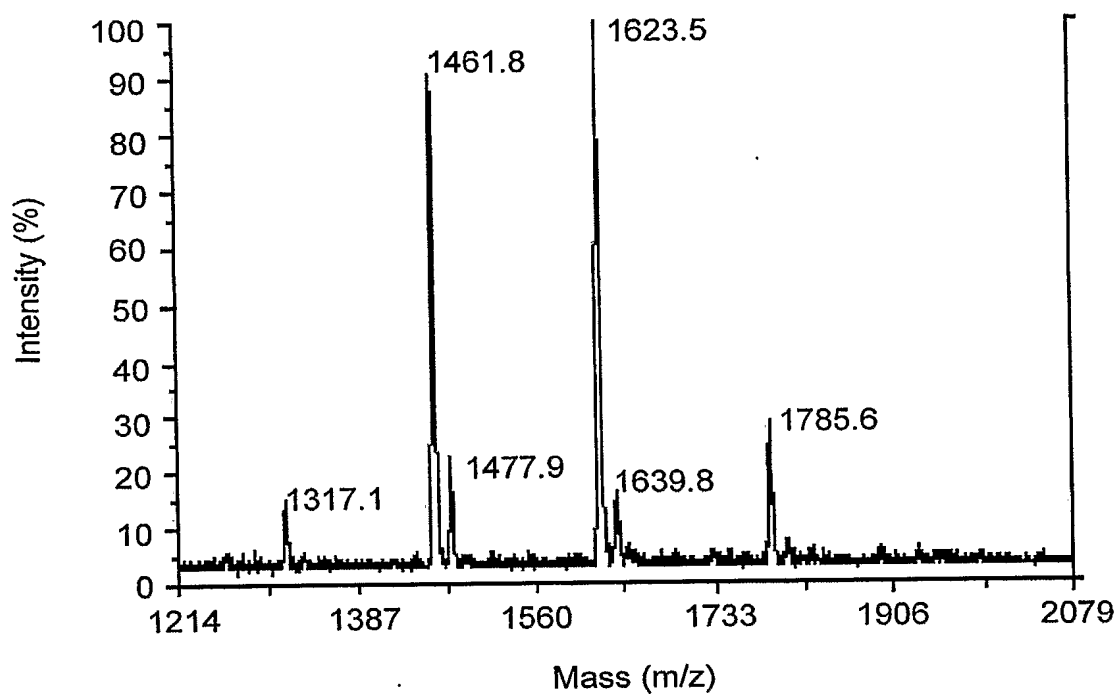
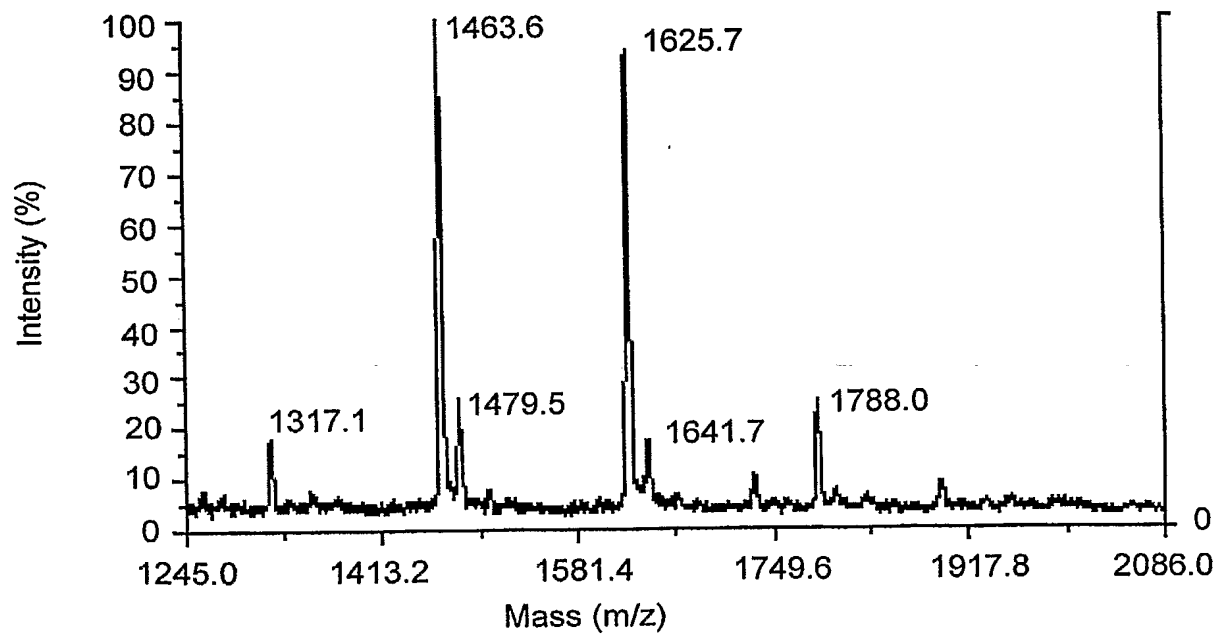


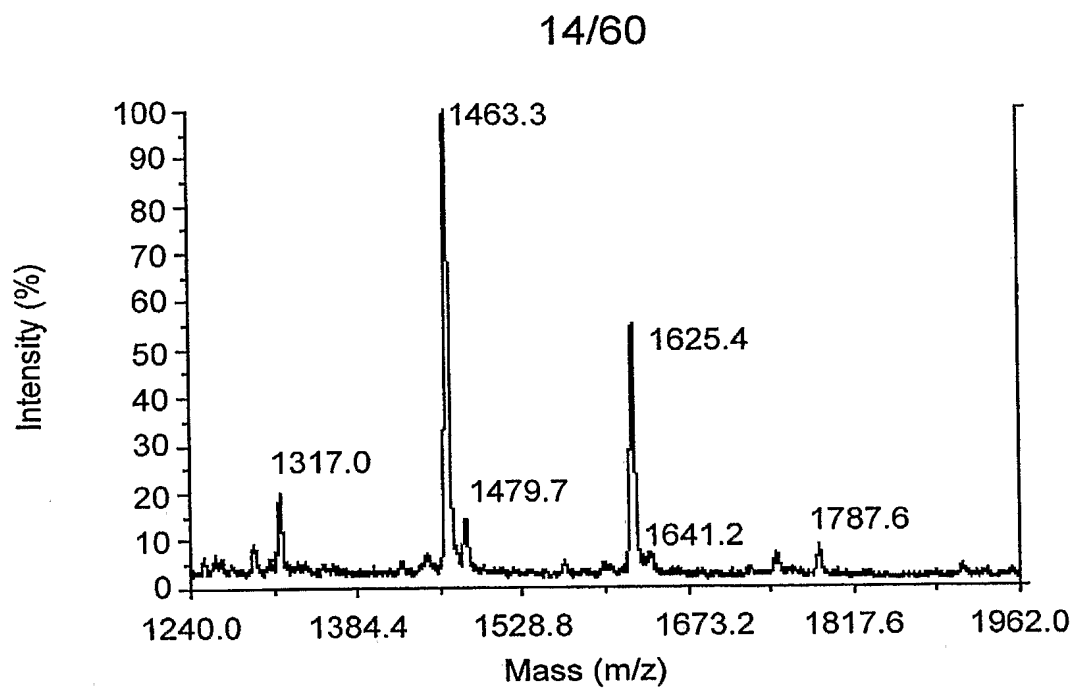
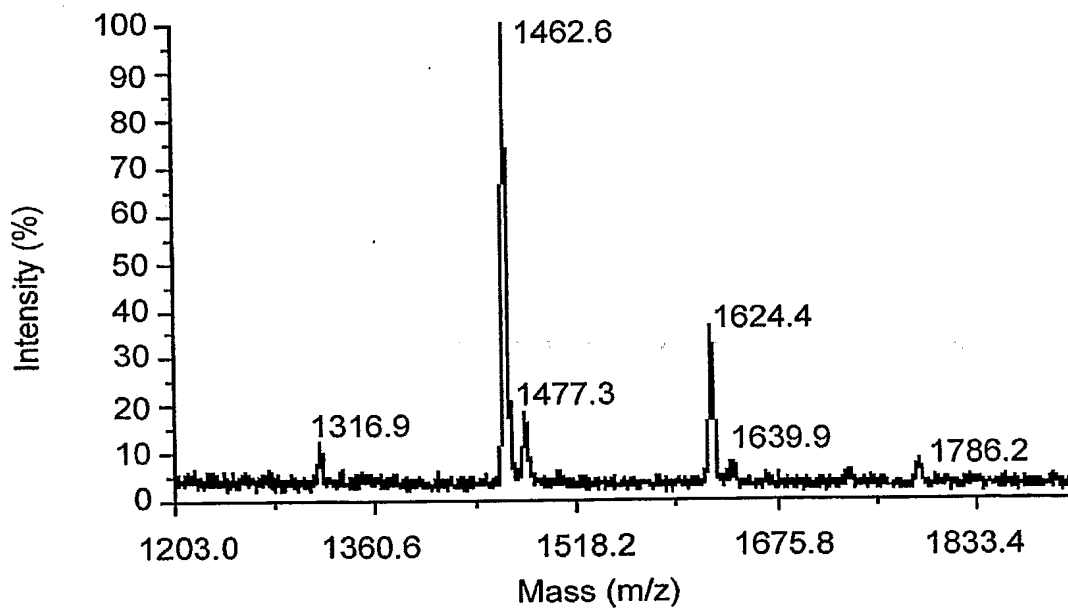
FIG. 5D

12/60

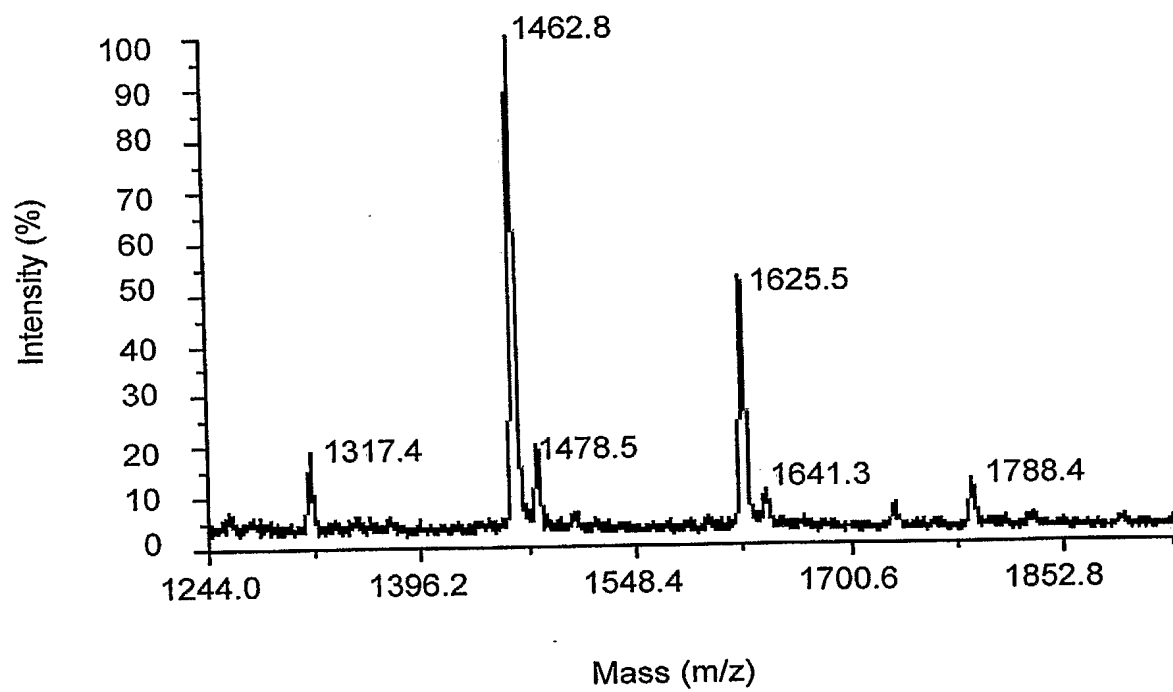
**FIG. 6**

13/60

*FIG. 7A**FIG. 7B*

*FIG. 7C**FIG. 7D*

15/60

*FIG. 7E*

16/60

GlcNAc β 1-2Man α 1 — 6 1317.2 Da
 Man β 1-4GlcNAc β 1-4GlcNAc
 GlcNAc β 1-2Man α 1 — 3

Fuc α 1
 |
 6
 1463.2 Da
 GlcNAc β 1-2Man α 1 — 6
 Man β 1-4GlcNAc β 1-4GlcNAc
 GlcNAc β 1-2Man α 1 — 3

GlcNAc β 1-2Man α 1 — 6 1478.8 Da (Linkage of
 Terminal Gal Unknown)
 *
 *
 Gal β 1
 *
 *
 GlcNAc β 1-2Man α 1 — 3
 Man β 1-4GlcNAc β 1-4GlcNAc

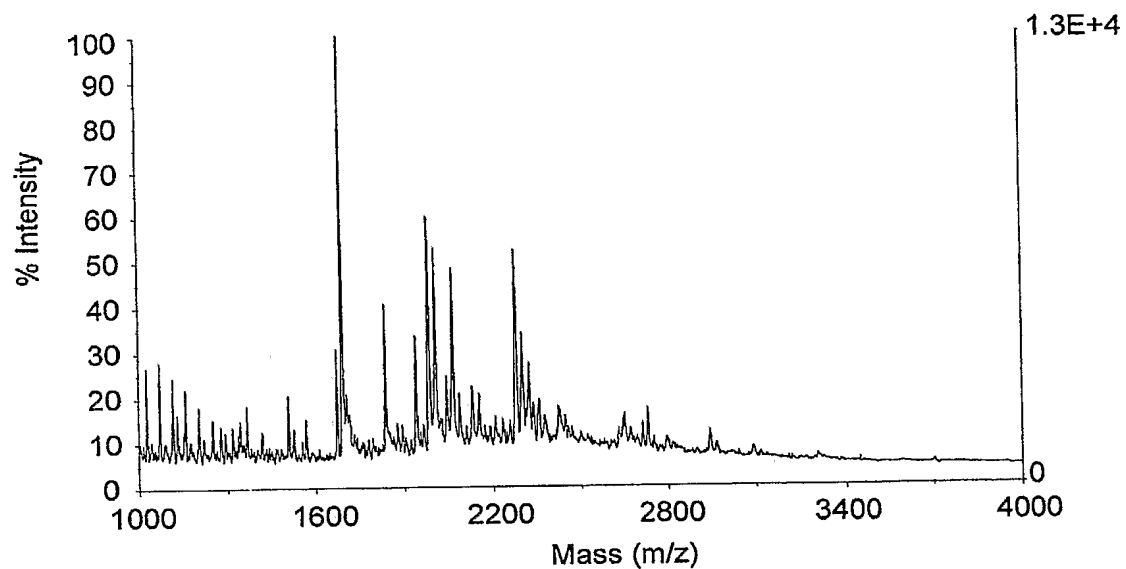
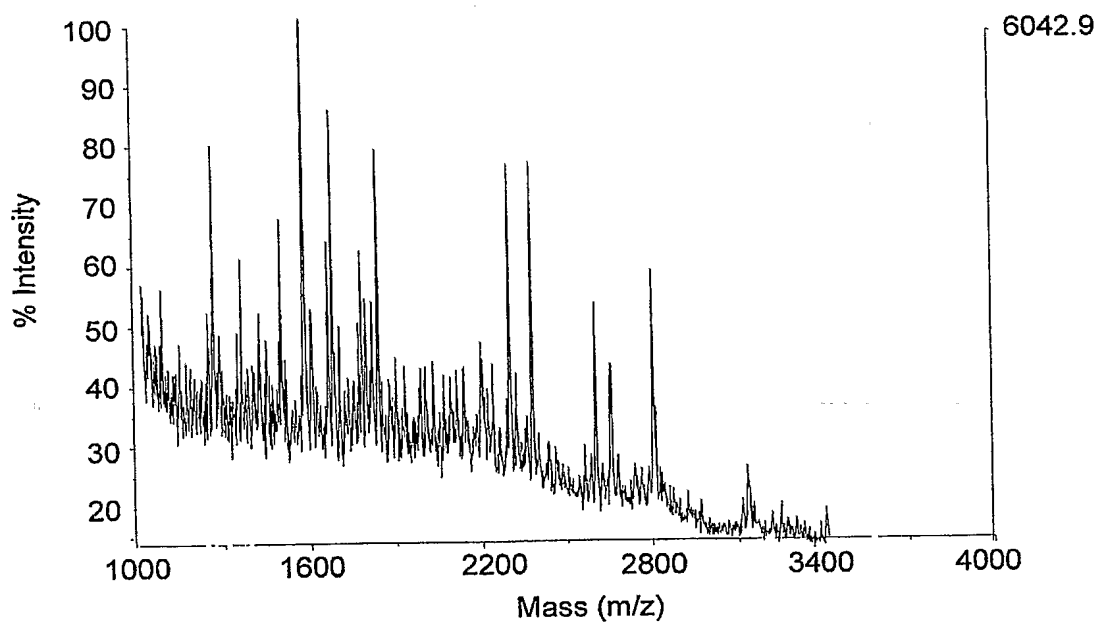
GlcNAc β 1-2Man α 1 — 6 1625.4 Da (Linkage of
 Terminal Gal Unknown)
 *
 *
 Gal β 1
 *
 *
 GlcNAc β 1-2Man α 1 — 3
 Man β 1-4GlcNAc β 1-4GlcNAc
 Fuc α 1
 |
 6

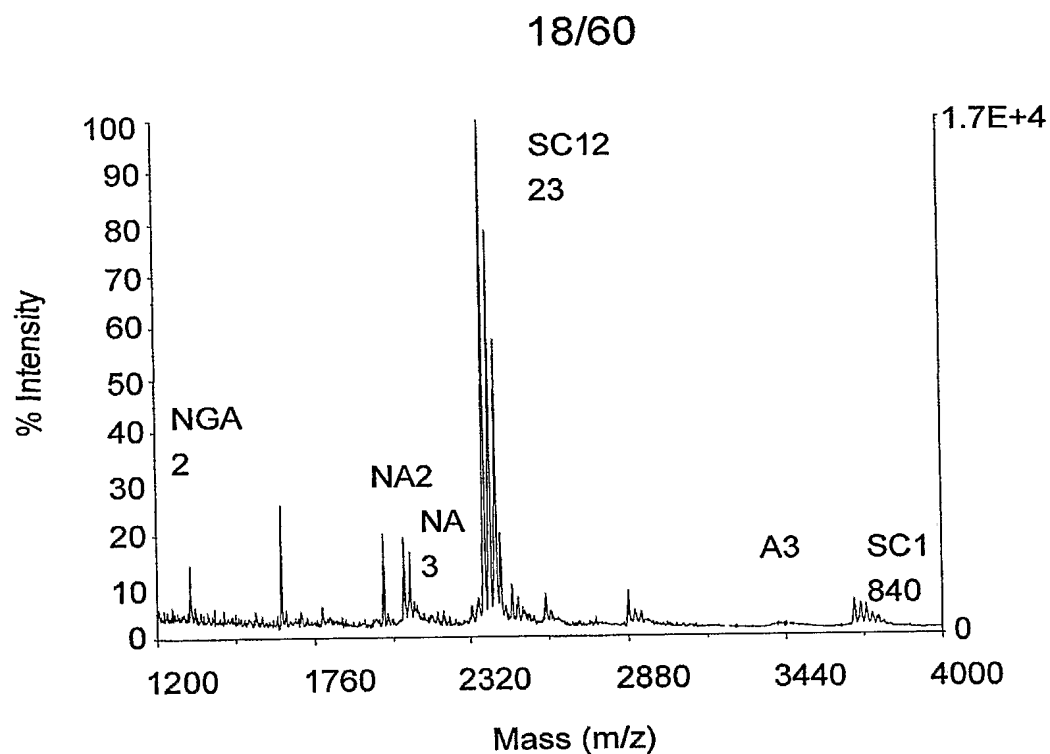
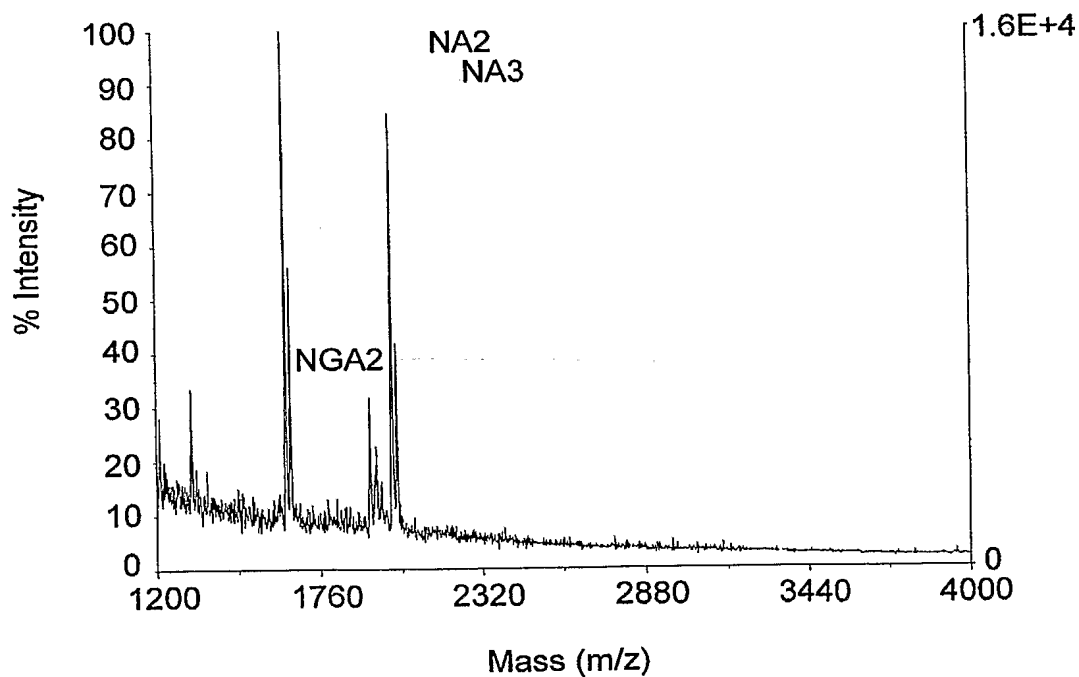
Gal β 1-4 GlcNAc β 1-2Man α 1 — 6 1640.9 Da
 Man β 1-4GlcNAc β 1-4GlcNAc
 Gal β 1-4 GlcNAc β 1-2Man α 1 — 3

Fuc α 1 1787.7 Da
 |
 6
 Gal β 1-4 GlcNAc β 1-2Man α 1 — 6
 Man β 1-4GlcNAc β 1-4GlcNAc
 Gal β 1-4 GlcNAc β 1-2Man α 1 — 3

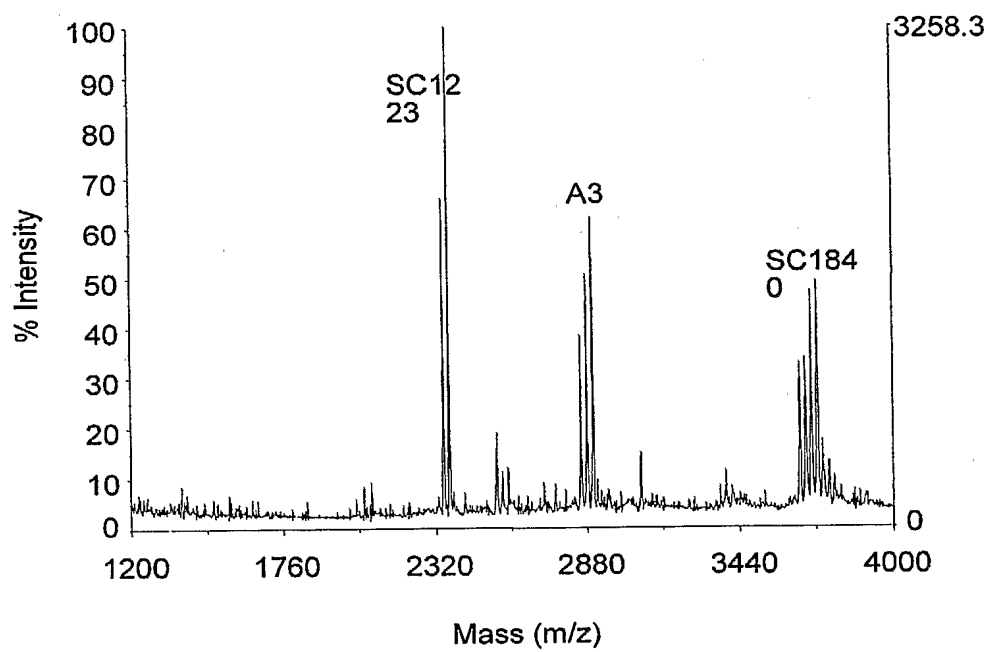
FIG. 8

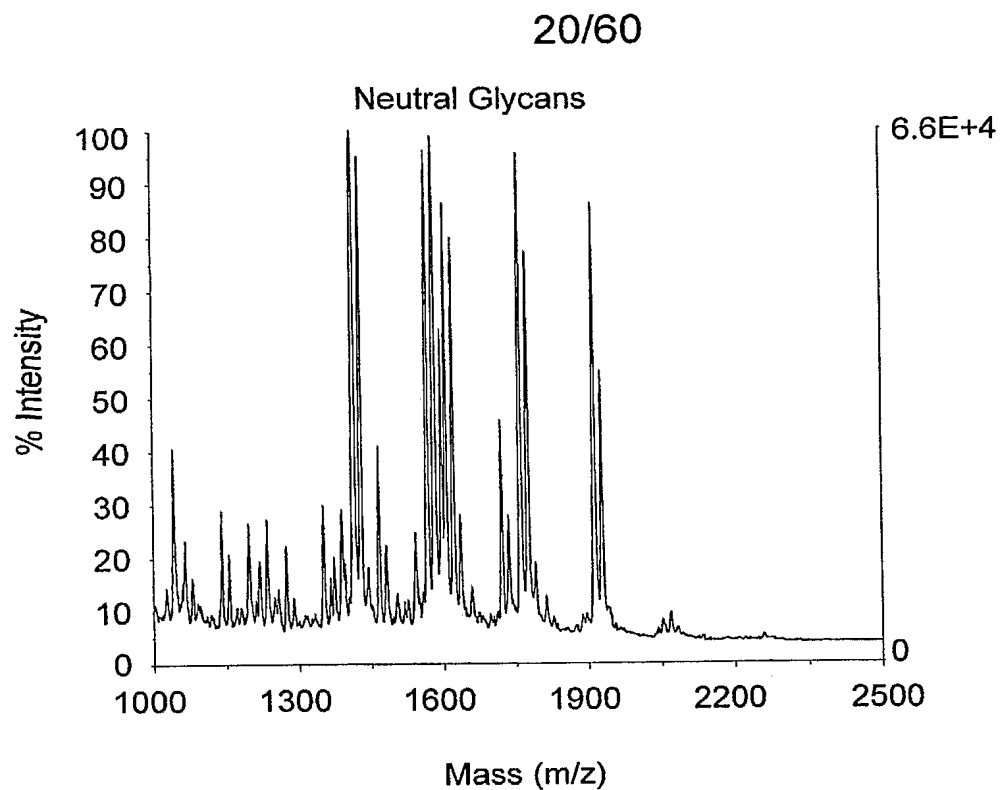
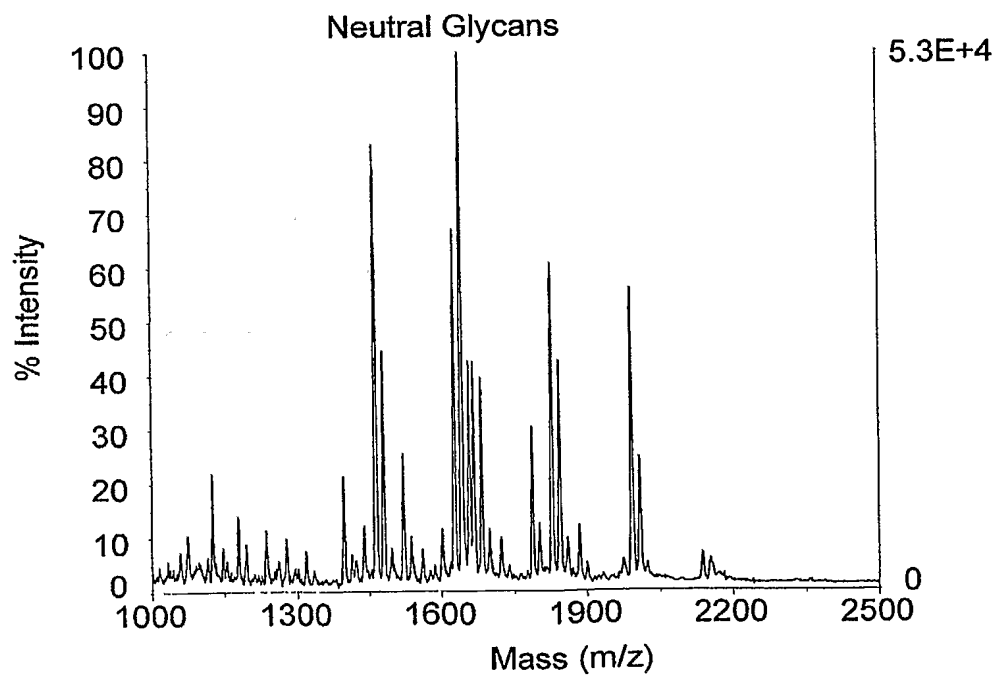
17/60

*FIG. 9A**FIG. 9B*

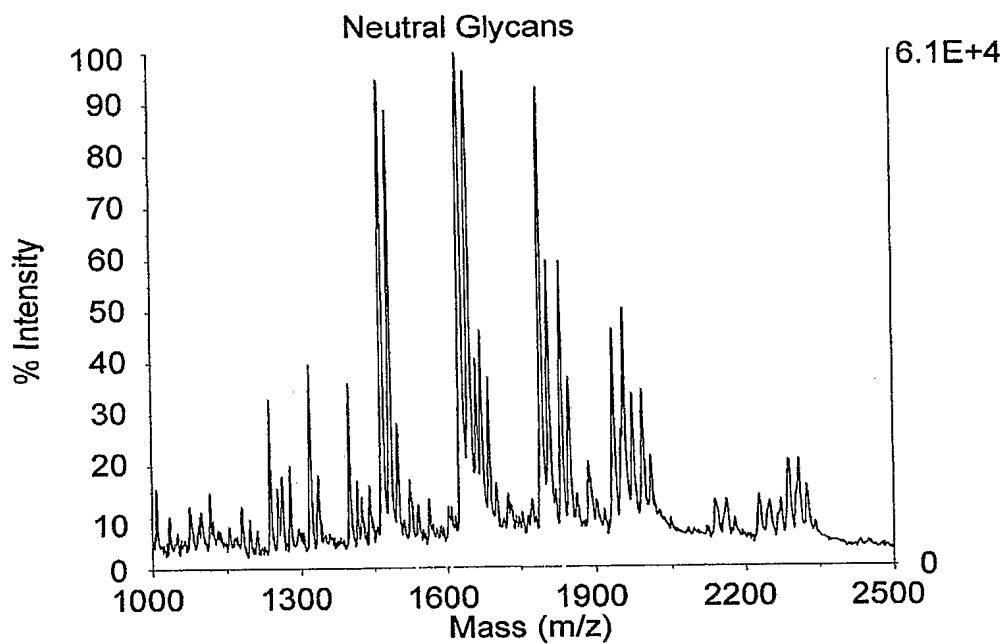
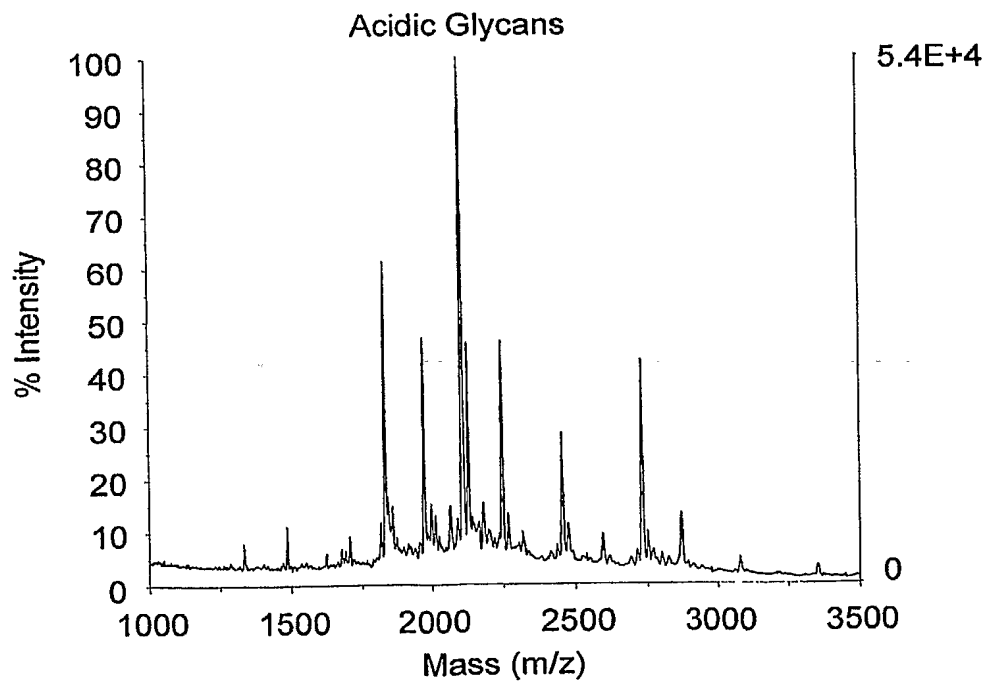
**FIG. 10A****FIG. 10B**

19/60

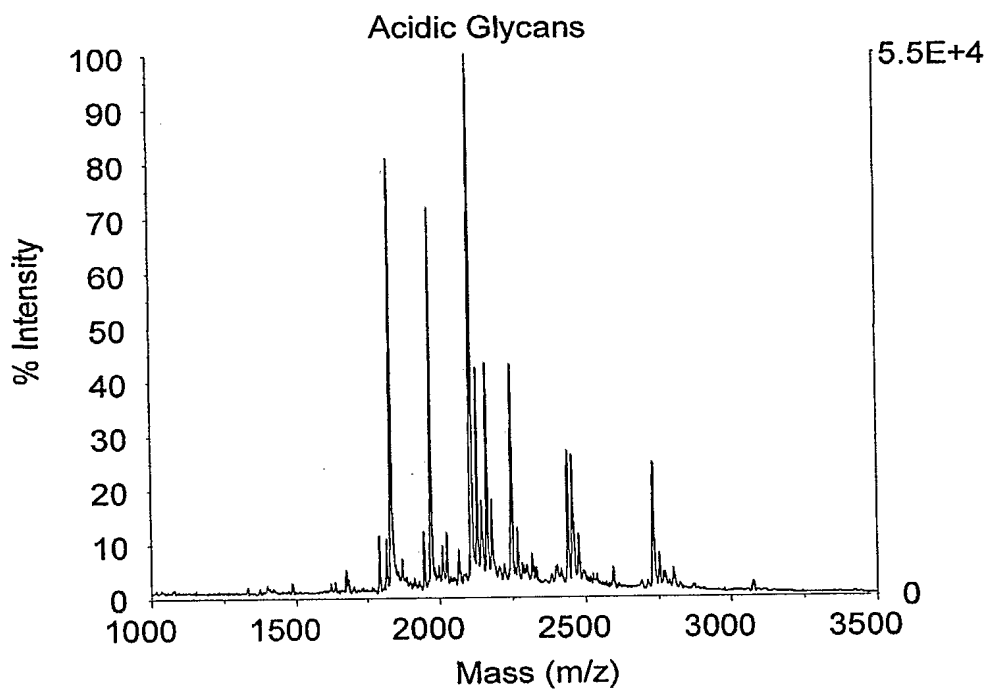
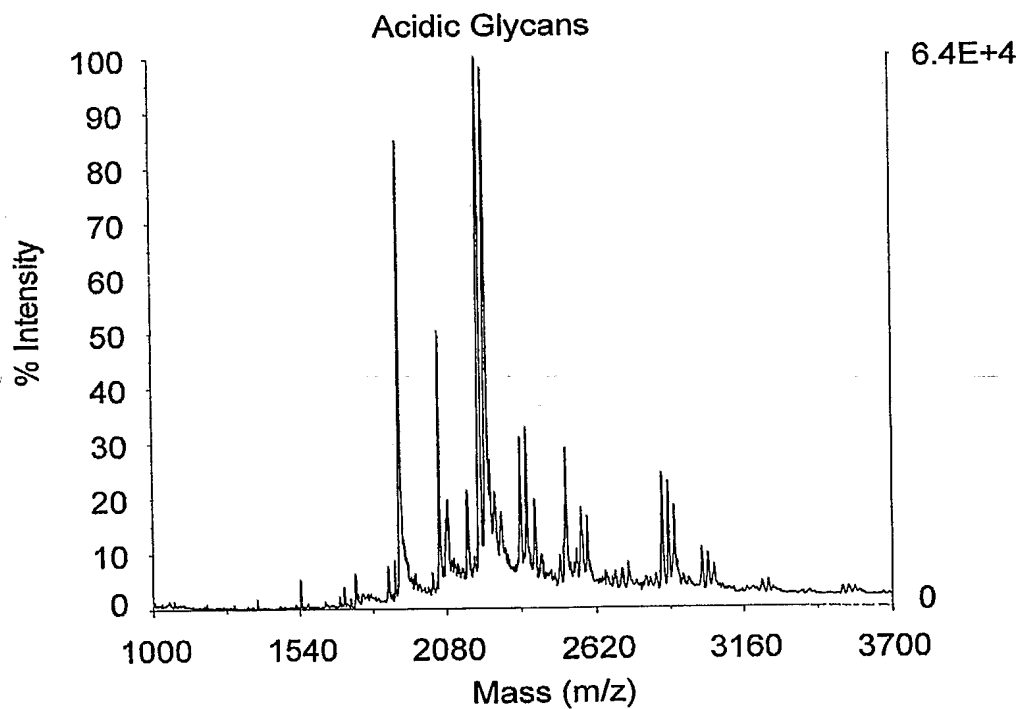
*FIG. 10C*

*FIG. 11A**FIG. 11B*

21/60

*FIG. 11C**FIG. 11D*

22/60

**FIG. 11E****FIG. 11F**

23/60

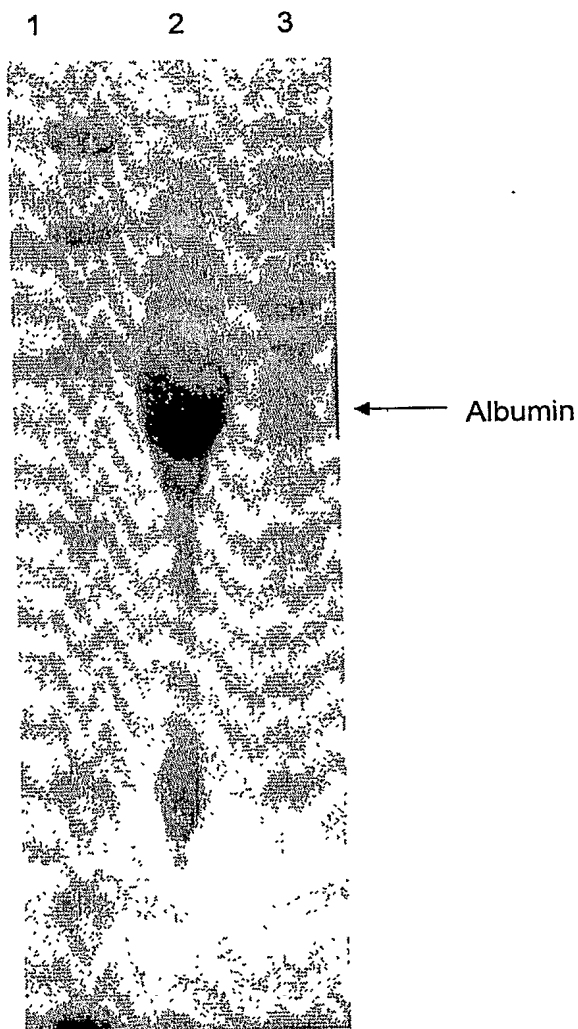
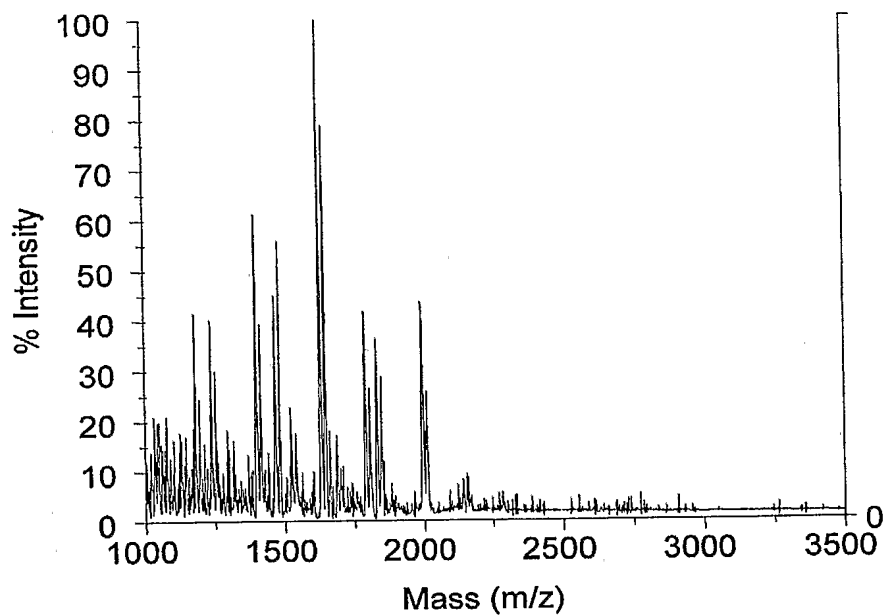
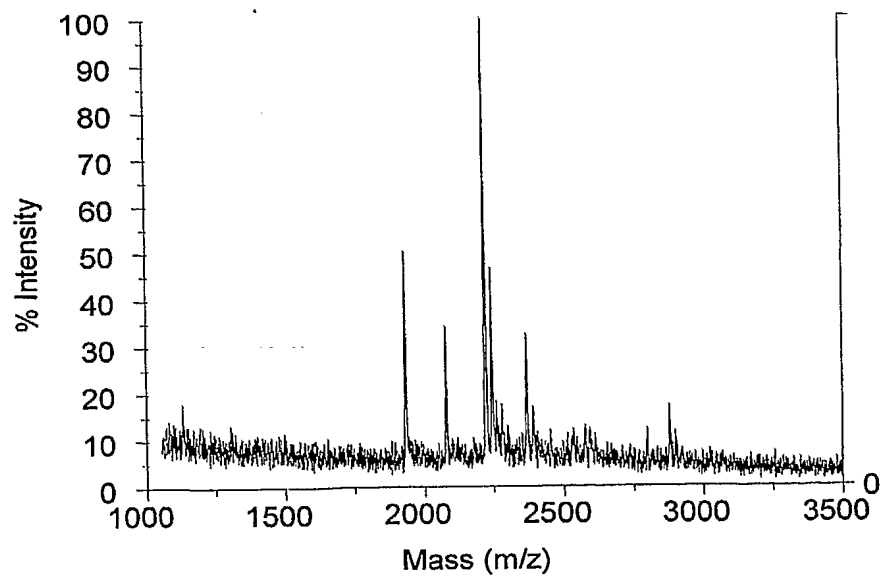


FIG. 12A

24/60

*FIG. 12B**FIG. 12C*

25/60

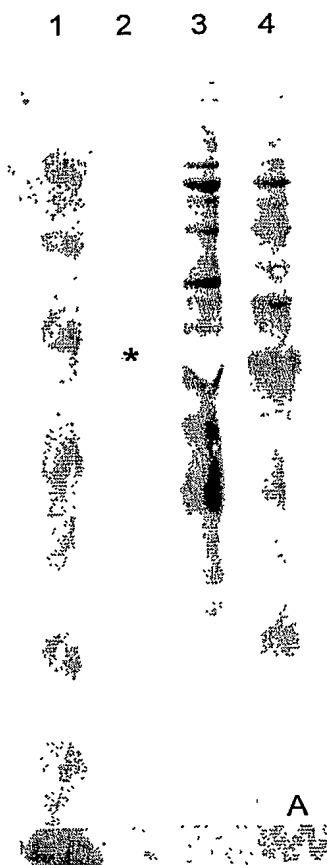
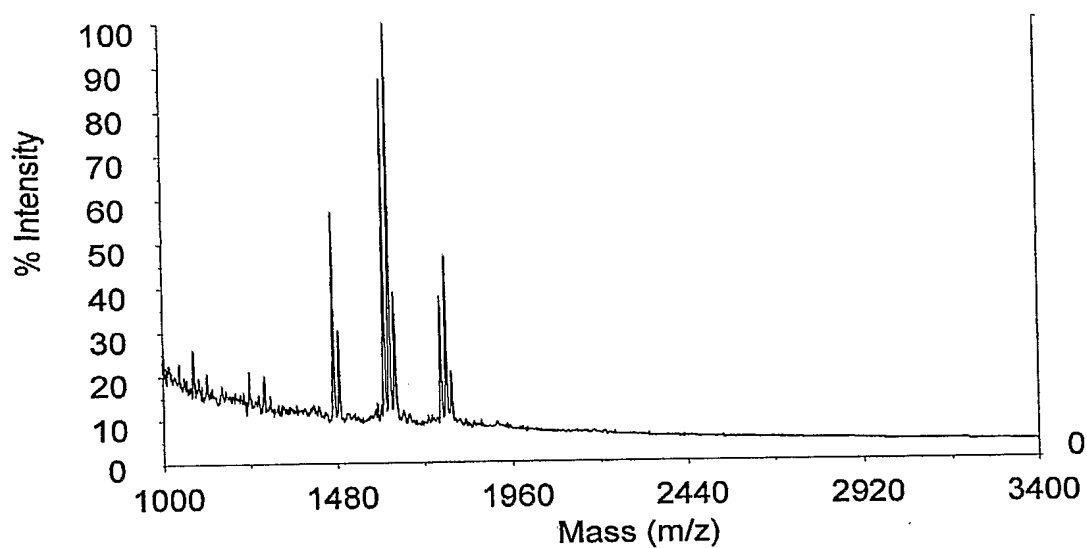
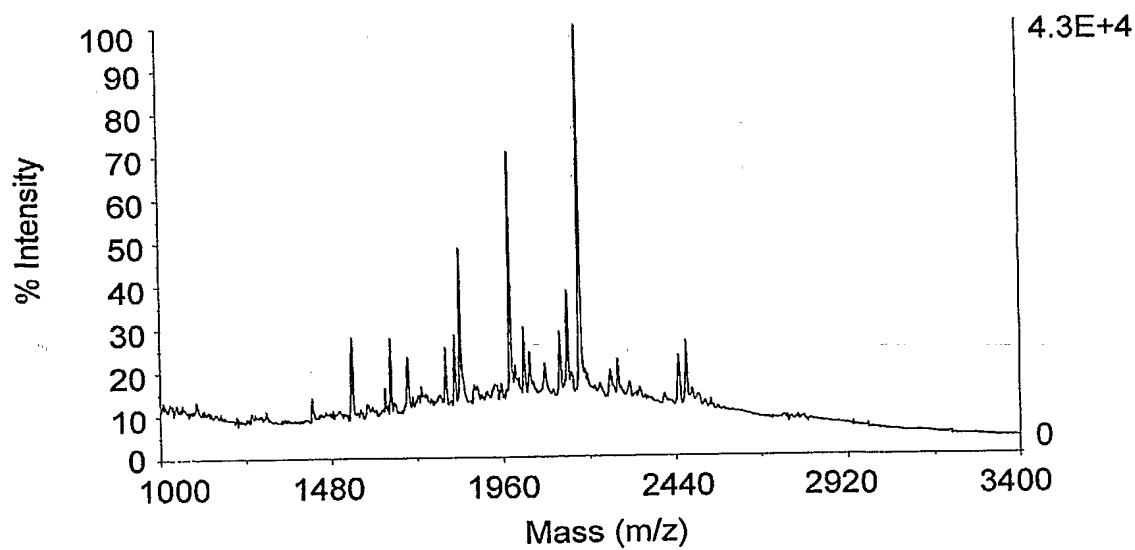
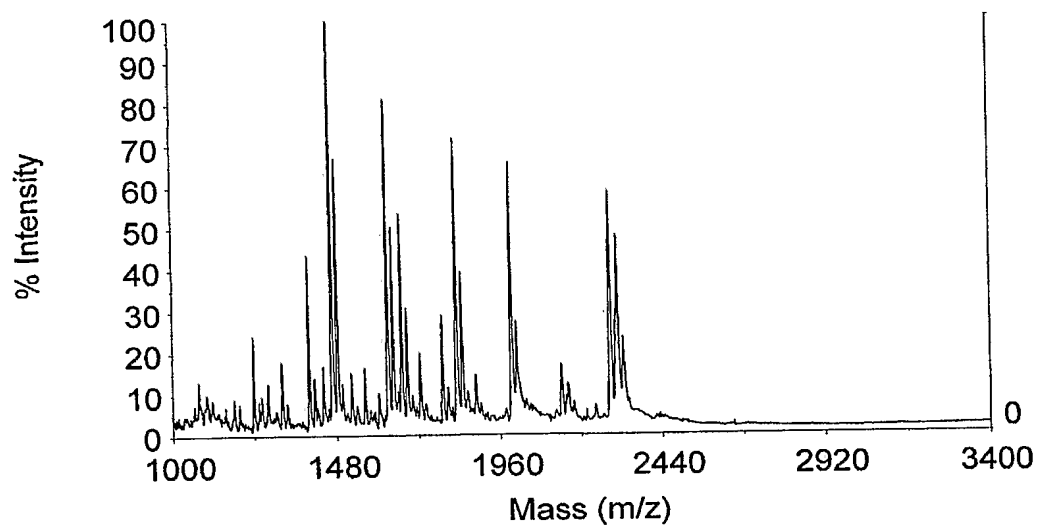
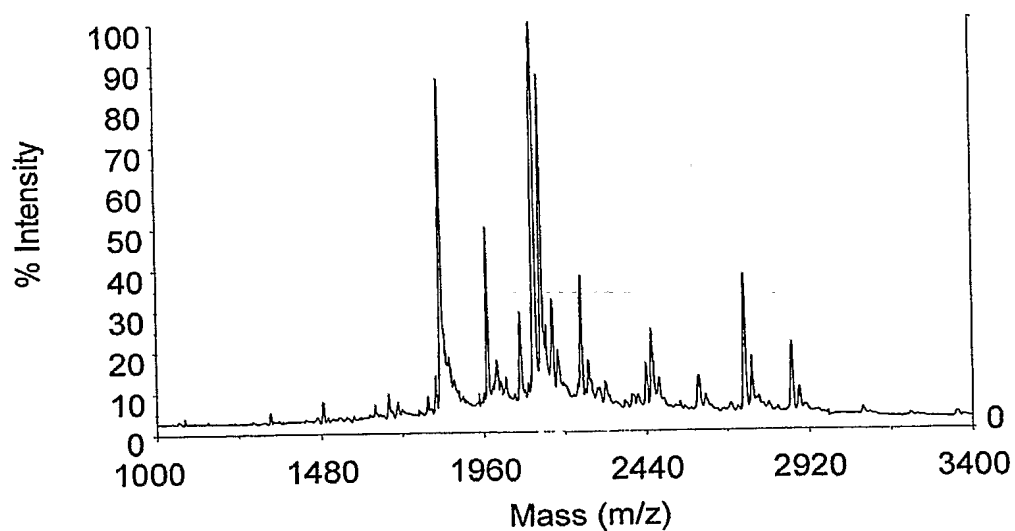


FIG. 13A

26/60

*FIG. 13B**FIG. 13C*

27/60

*FIG. 13D**FIG. 13E*

28/60

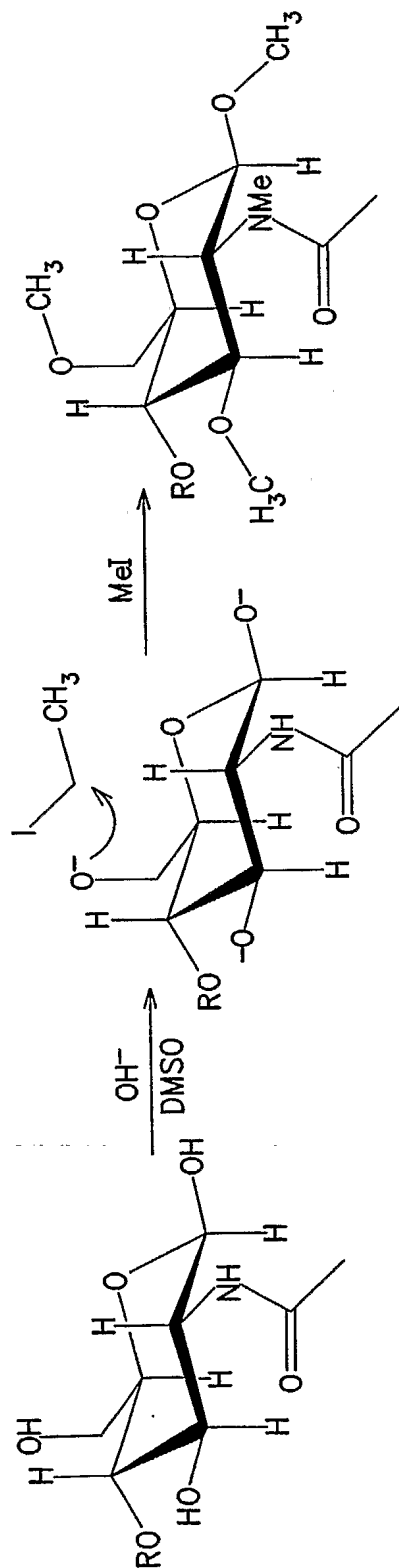
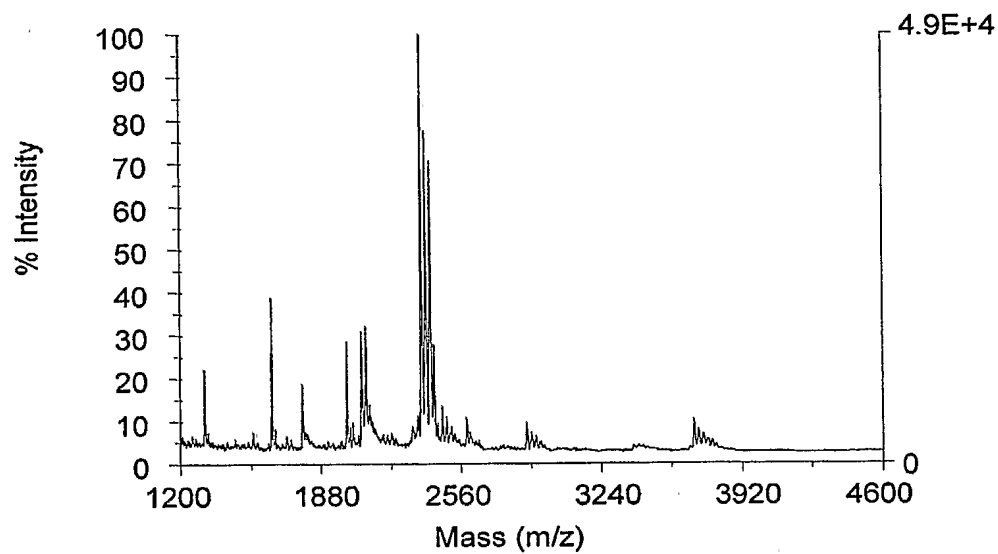
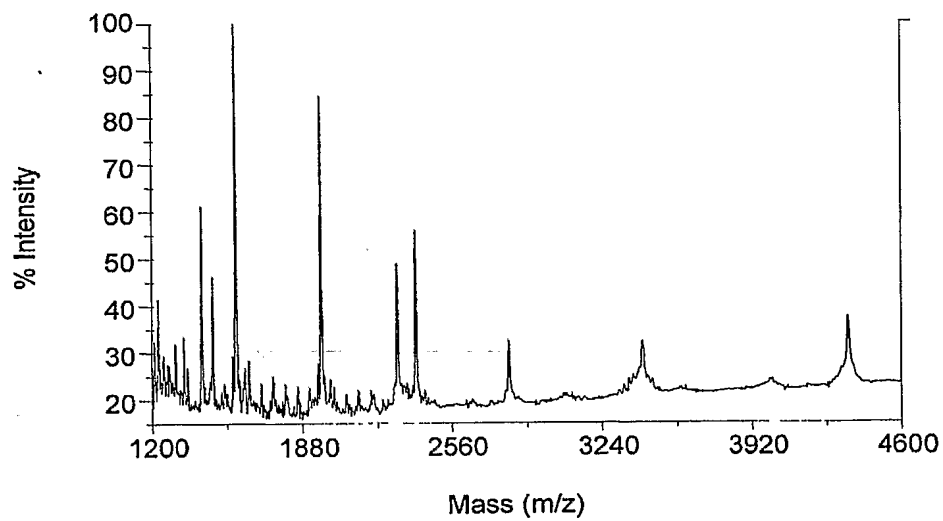


FIG. 14

29/60

*FIG. 15A**FIG. 15B*

30/60

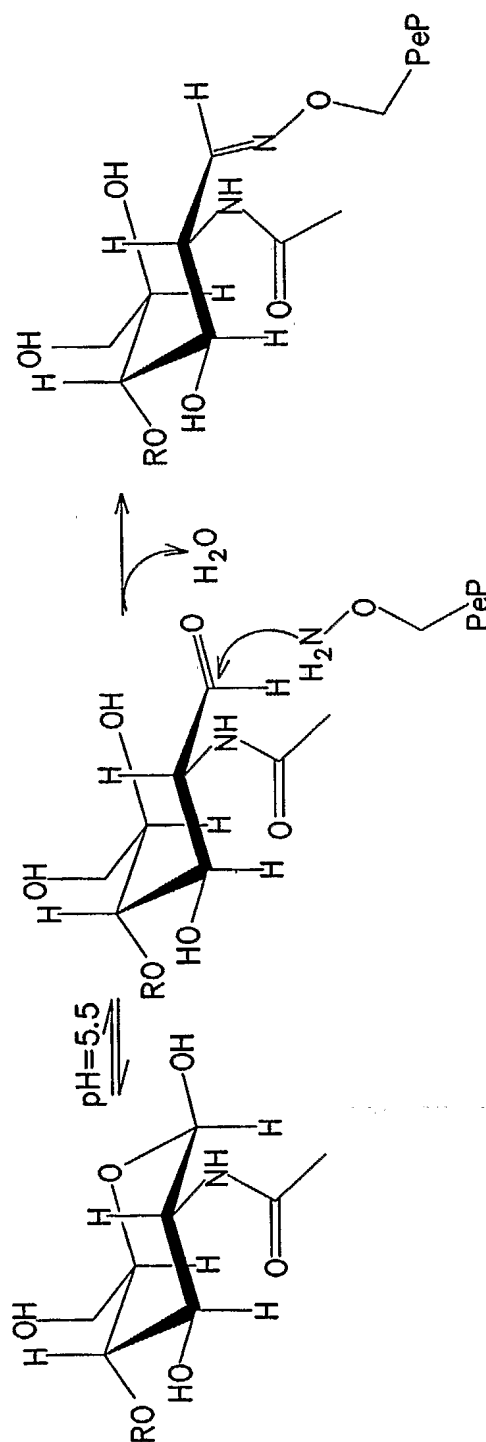
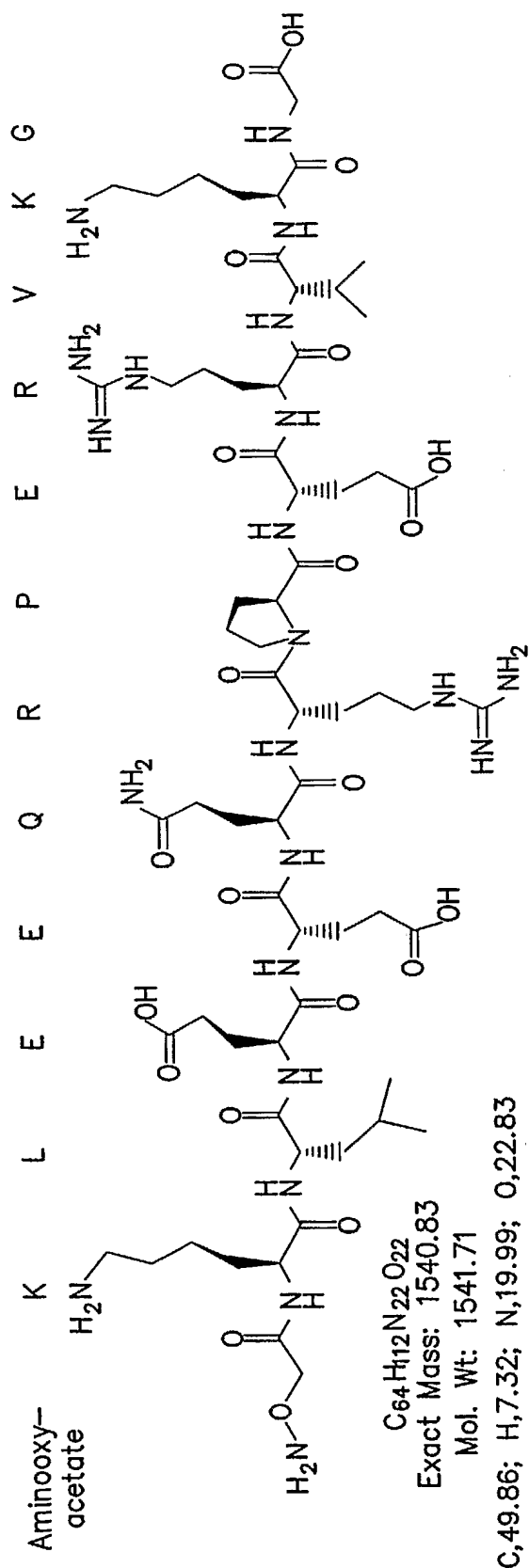
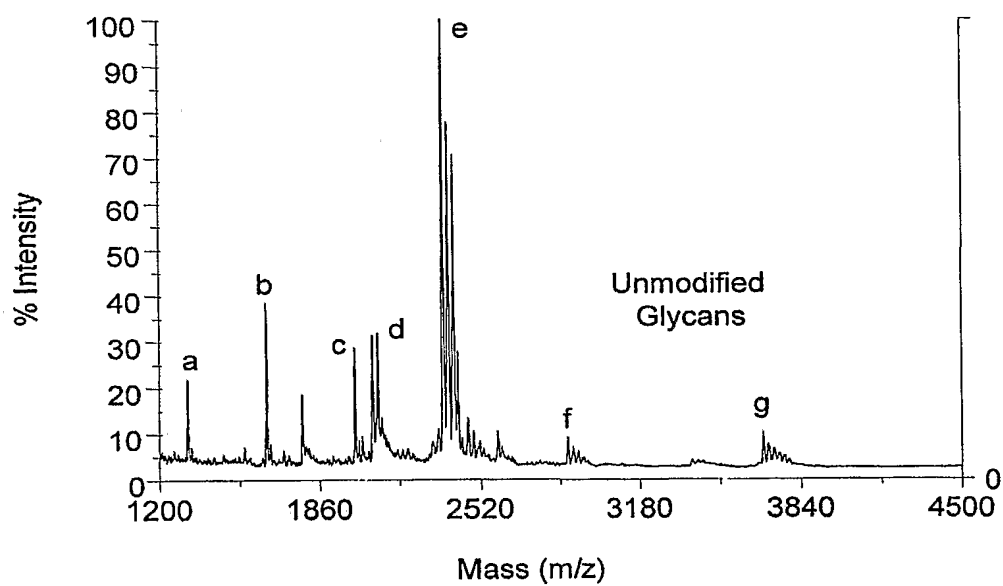
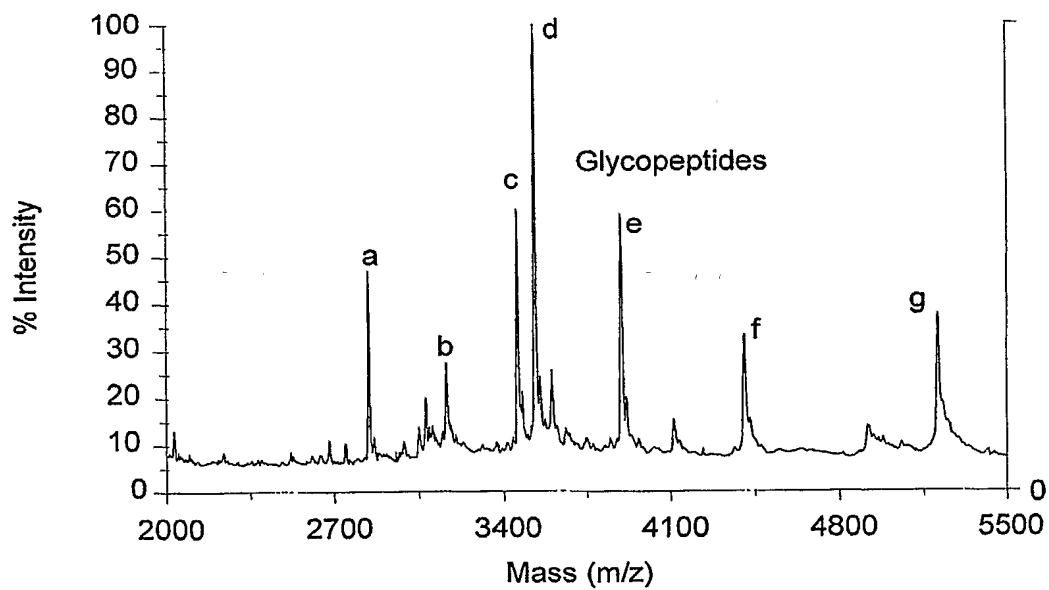
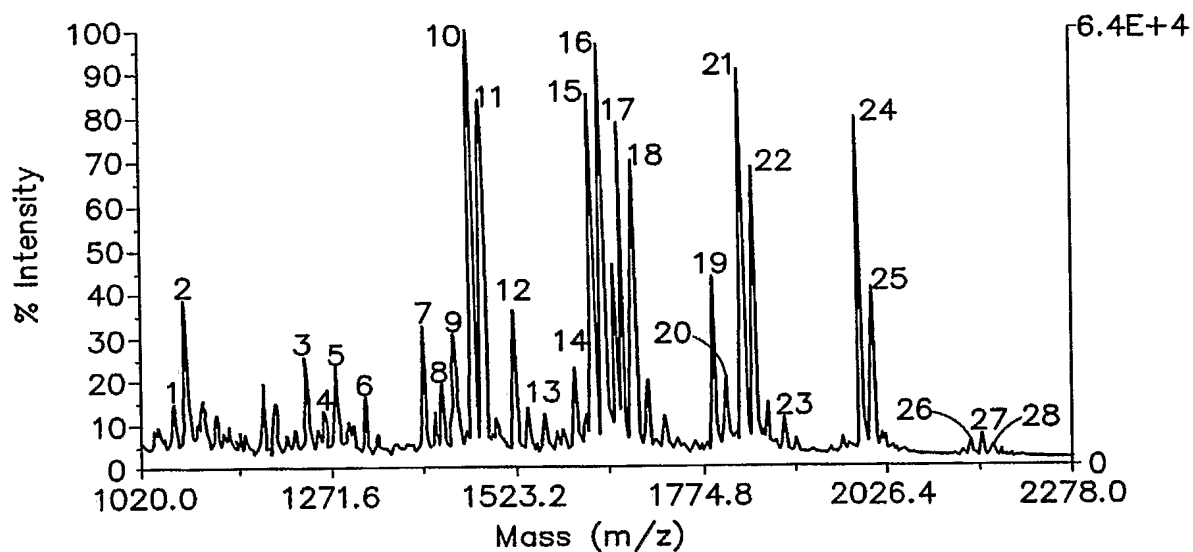
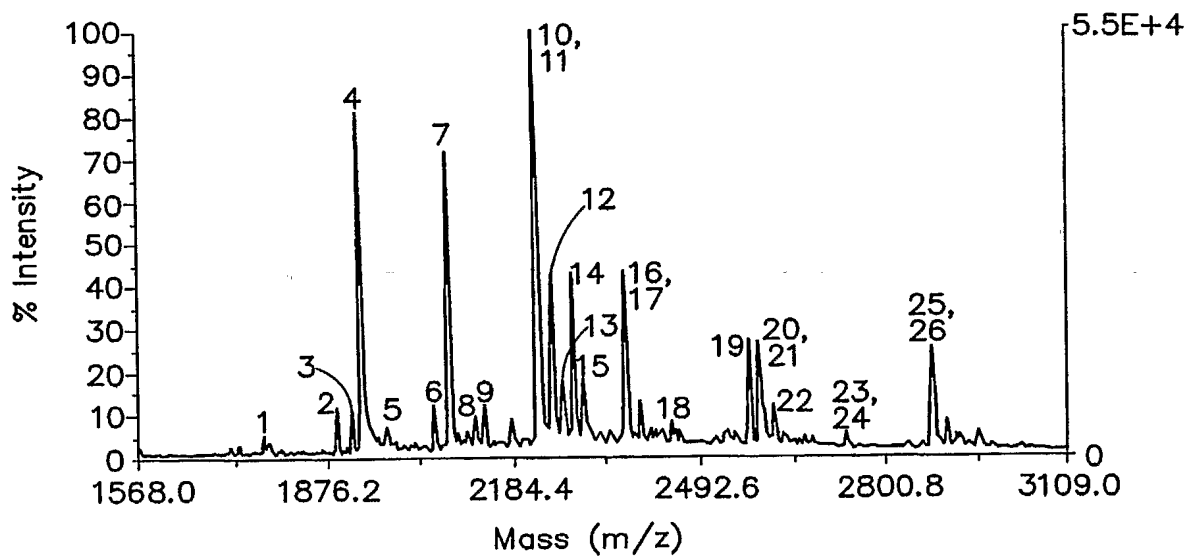


FIG. 16

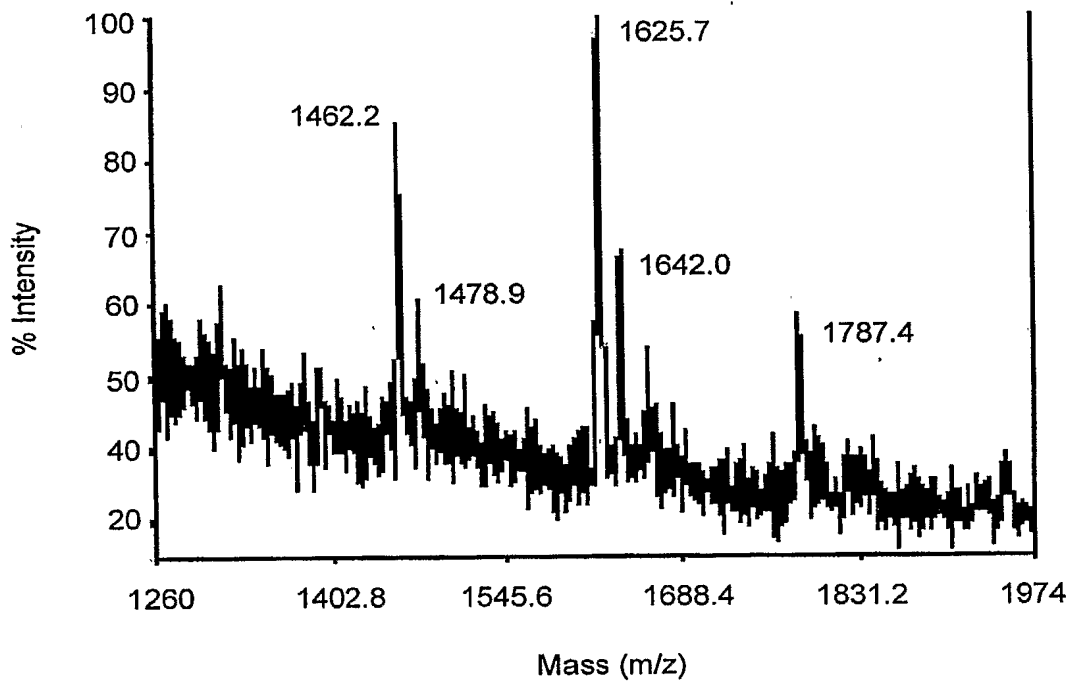
31/60

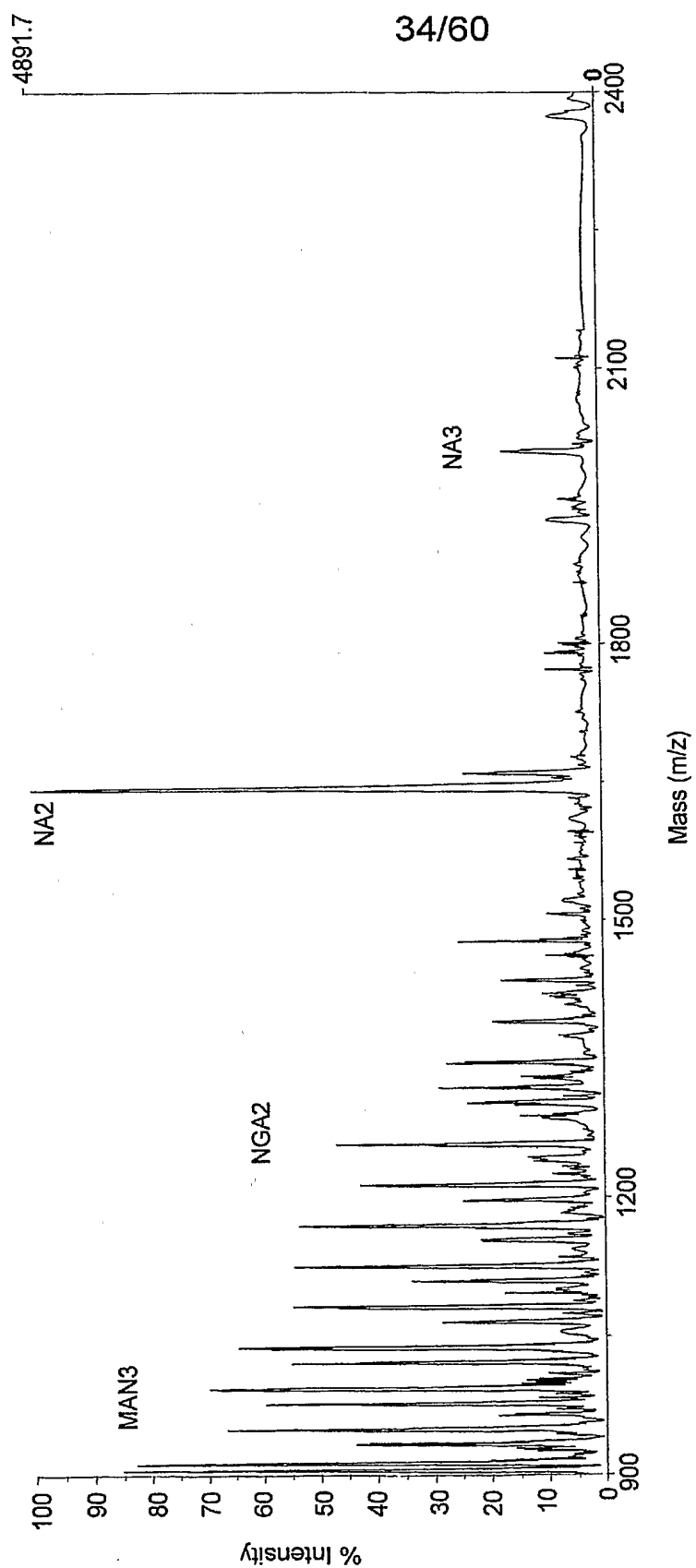
**FIG. 17A****FIG. 17B**

32/60

**FIG. 18A****FIG. 18B**

33/60

*FIG. 19*



DHB-Spermine Neutrals, 1 pmol

FIG. 20A

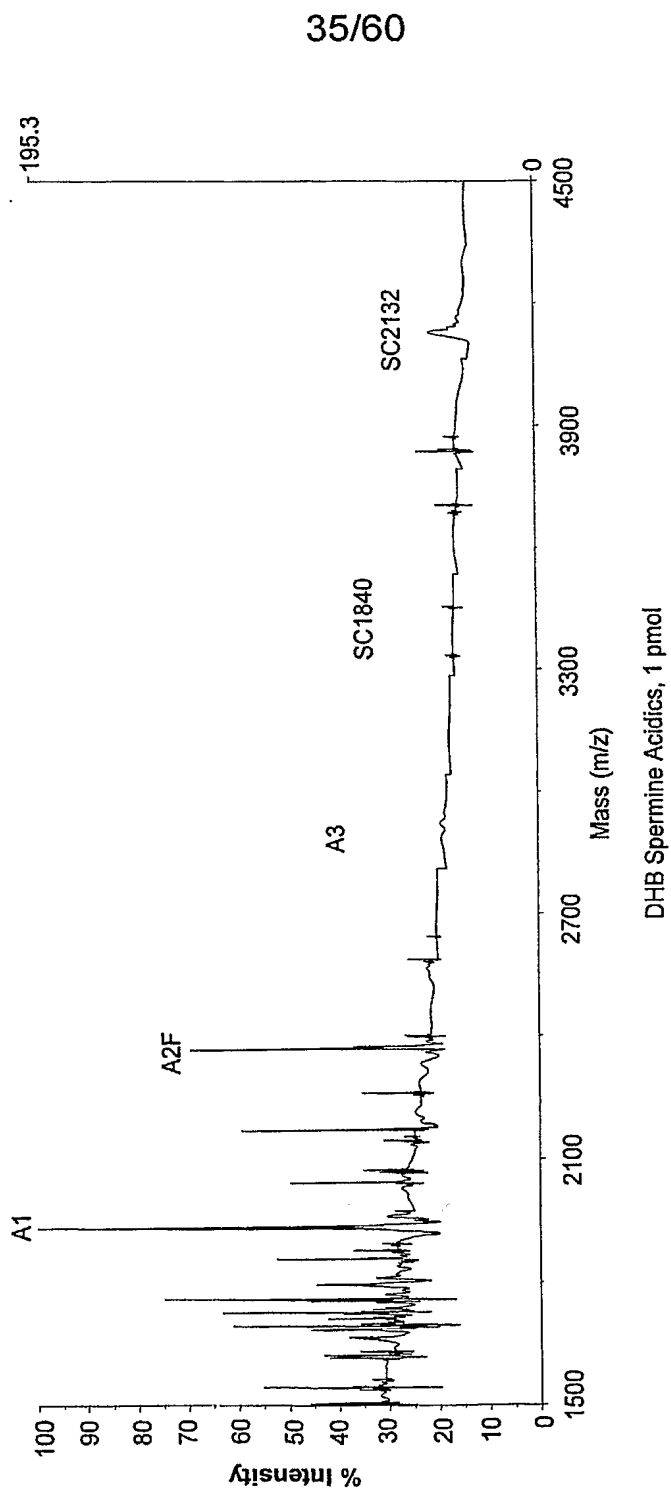
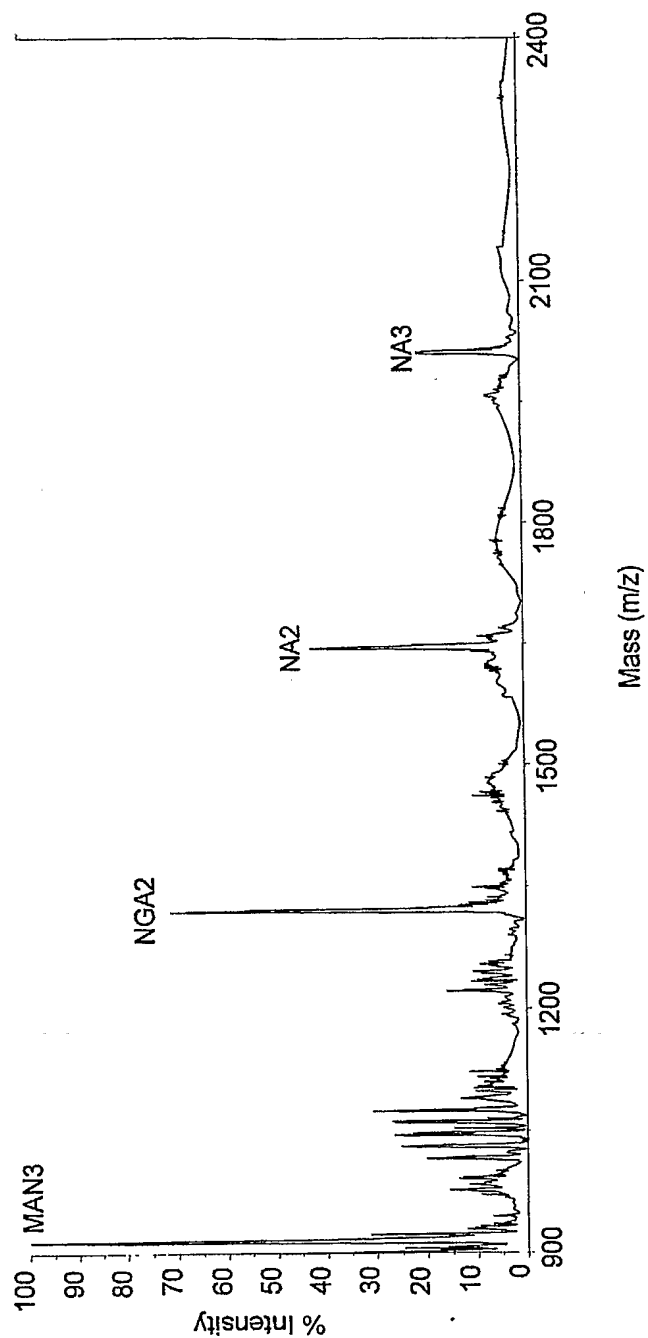


FIG. 20B

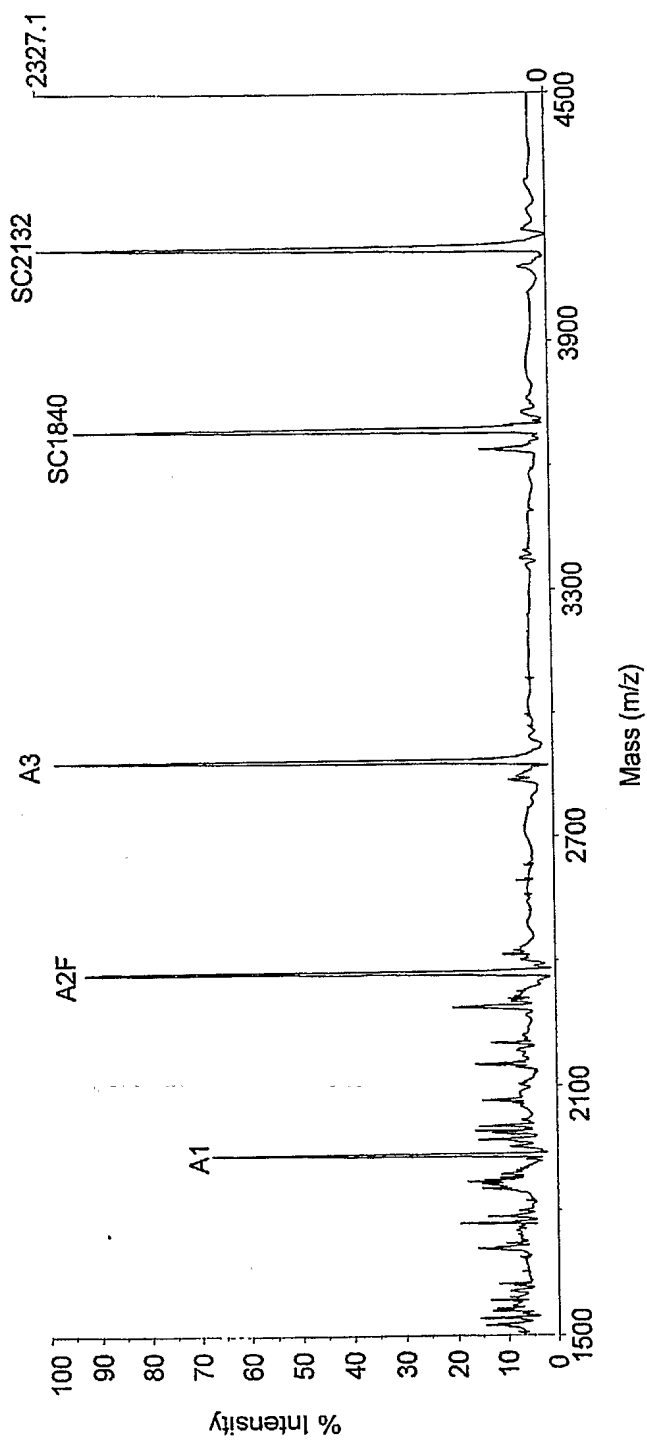
36/60



5-MSA/DHB Neutrals, 25 fmol

FIG. 20C

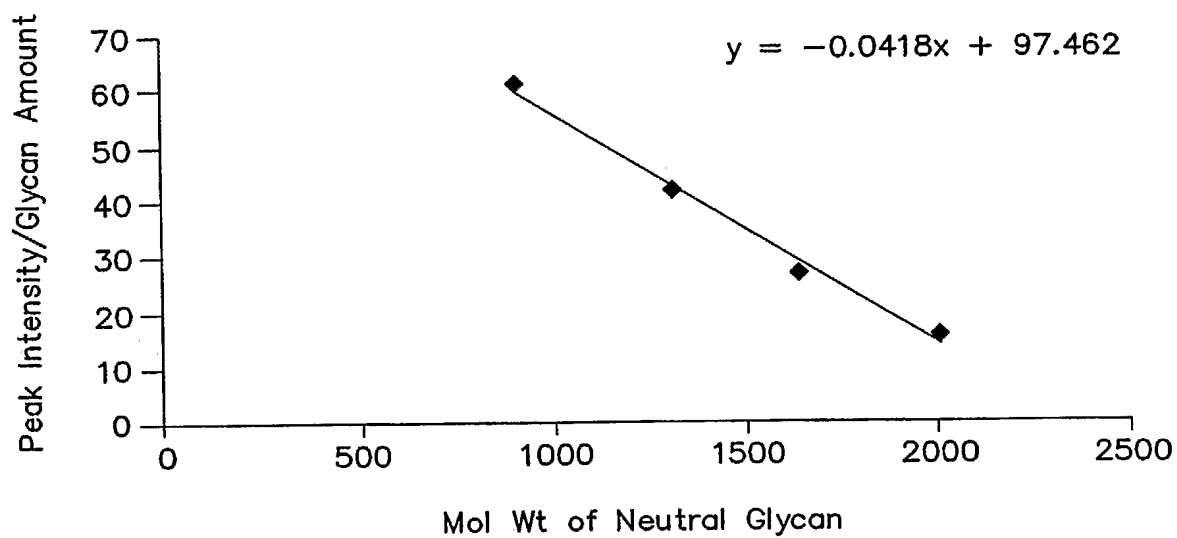
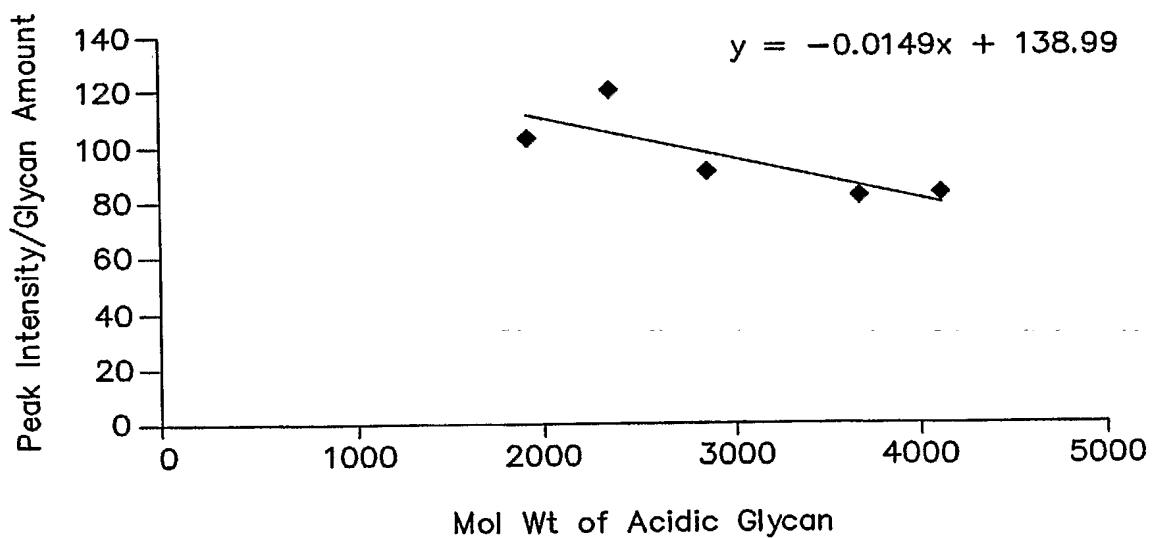
37/60



Nafion + ATT Acids, 25 fmol

FIG. 20D

38/60

**FIG. 21A****FIG. 21B**

39/60

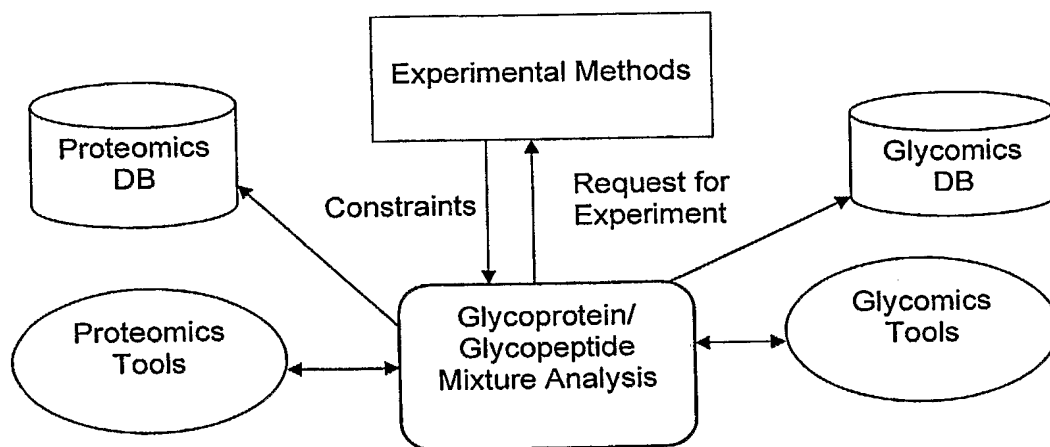
| Glycans in Spot | Man3 Intens. | NGA2 Intens. | NA2 Intens. | NGA3 Intens. |
|---------------------|--------------|--------------|-------------|--------------|
| Man3 | 5.9e3 | — | — | — |
| Man3 NGA2 | 6.5e3 | 4.4e3 | — | — |
| Man3 NGA2 NA2 | 6.4e3 | 4.5e3 | 2.0e3 | — |
| Man3 NGA2 NA2 | 8.4e3 | 4.8e3 | 2.3e3 | — |

FIG. 21C

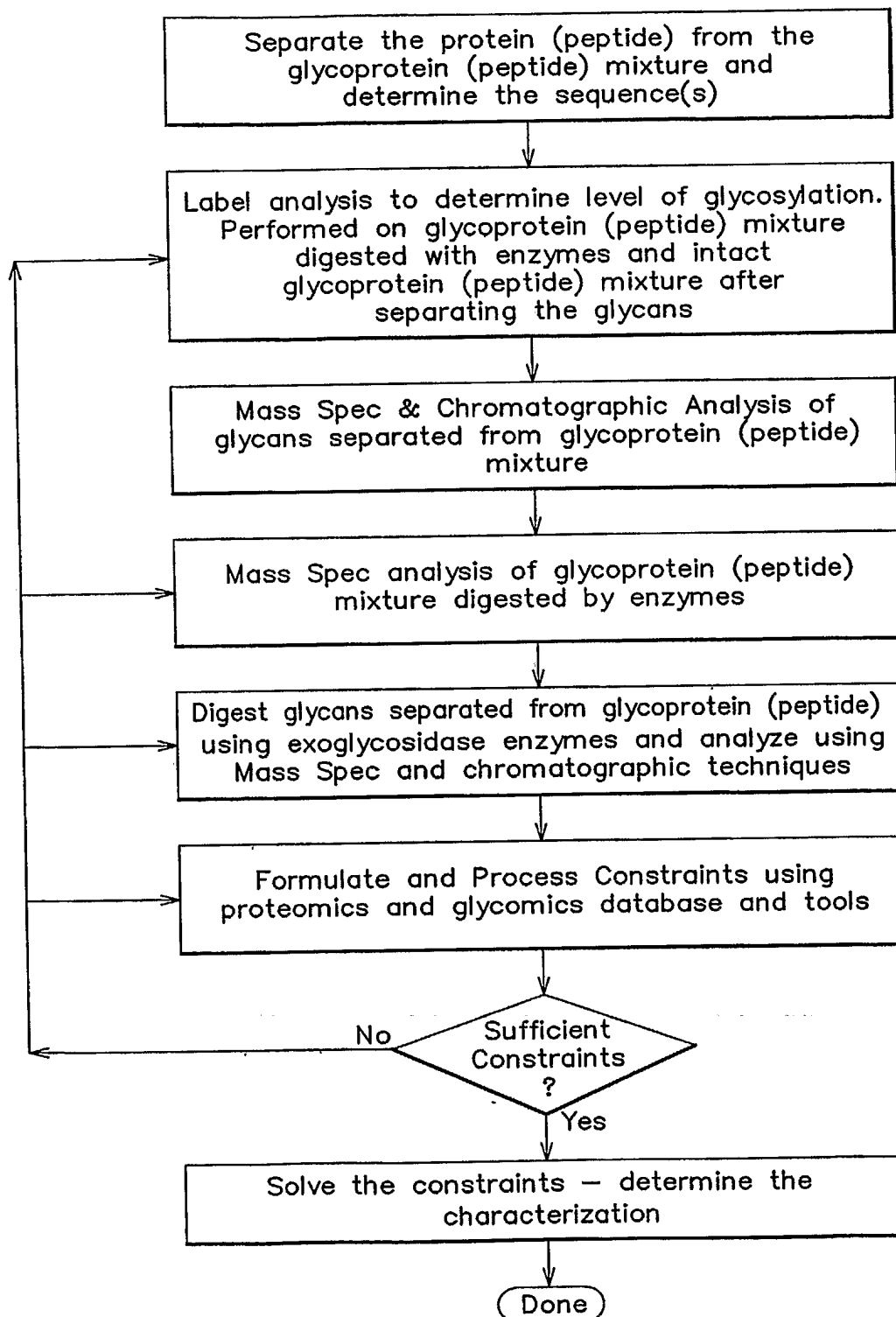
| Glycans in Spot | A1 Intens. | A2F Intens. | A3 Intens. | SC1840 Intens. | SC2132 Intens. |
|-------------------------------------|------------|-------------|------------|----------------|----------------|
| A1 | 2.2e3 | — | — | — | — |
| A1 A2F | 4.6e3 | 2.3e3 | — | — | — |
| A1 A2F A3 | 2.6e3 | 1.2e3 | 3.6e3 | — | — |
| A1 A2F A3 SC1840 | 3.2e3 | 2.0e3 | 5.1e3 | 3.5e3 | — |
| A1 A2F A3 SC1840 SC2132 | 2.8e3 | 1.6e3 | 3.9e3 | 2.0e3 | 2.5e3 |

FIG. 21D

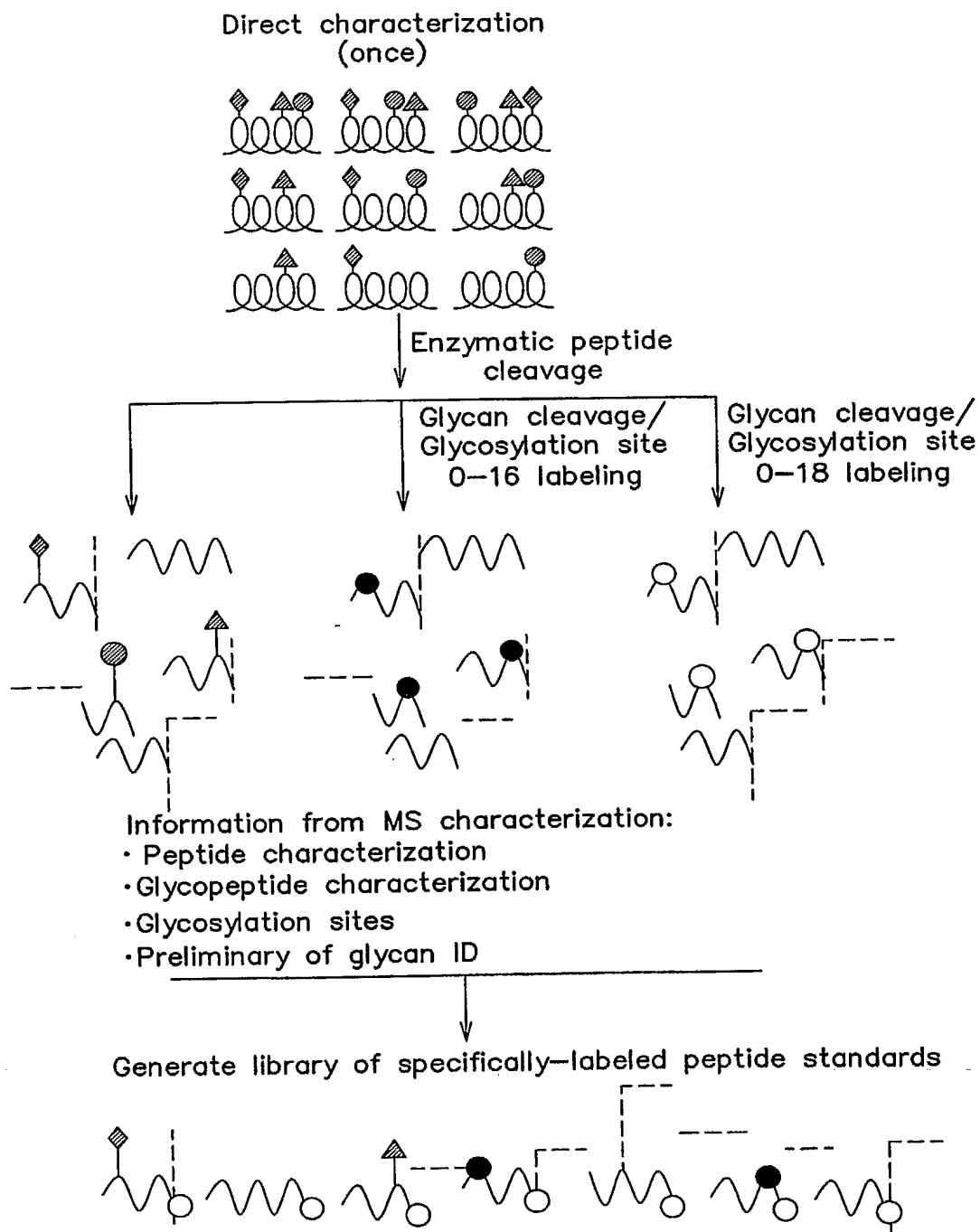
40/60

*FIG. 22*

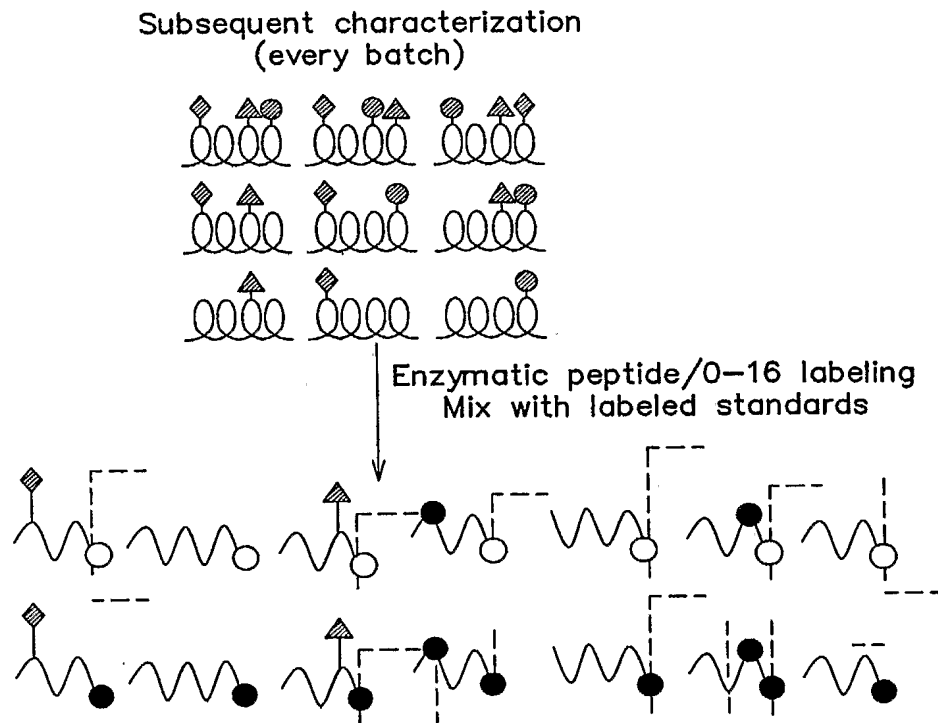
41/60

**FIG. 23**

42/60

**FIG. 24A**

43/60

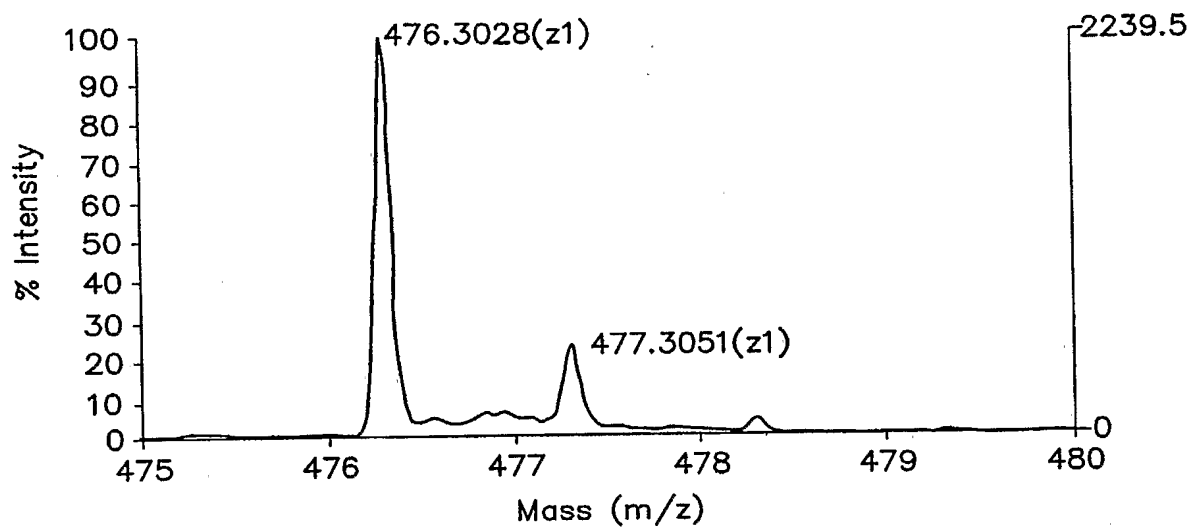
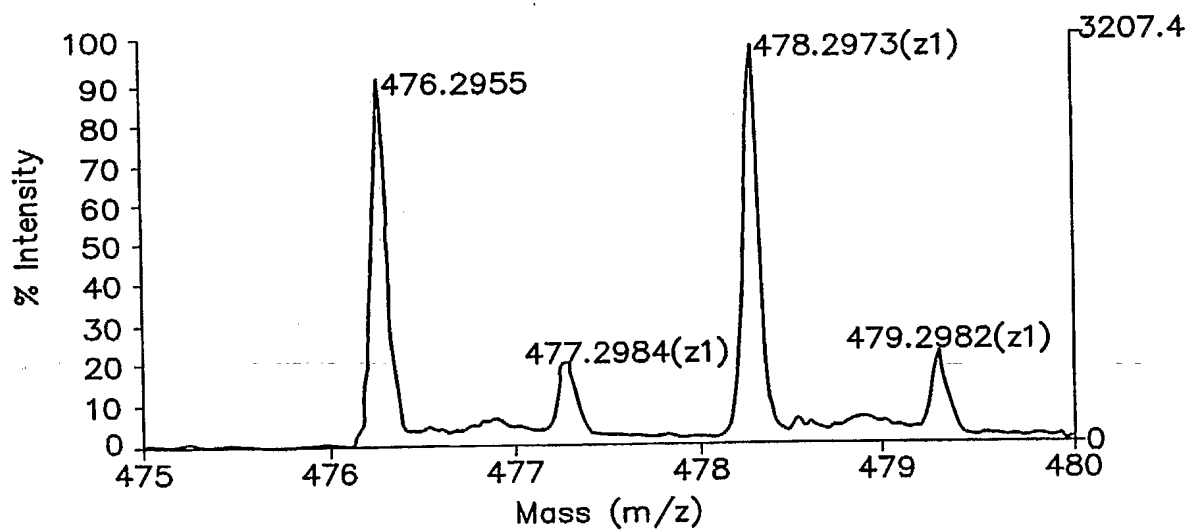


Information from MS characterization:

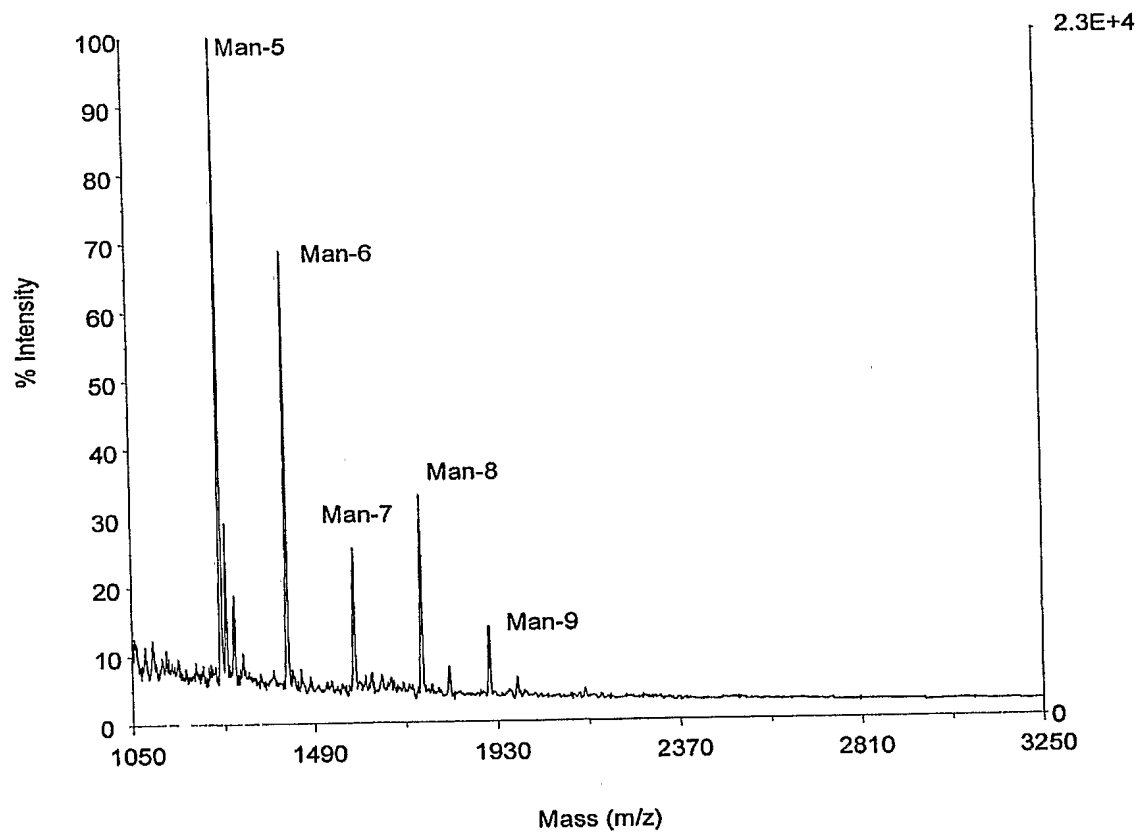
- Quantitative glycan occupancy at each site
- Quantity of each different glycan at each site
- Preliminary ID of glycan structure at each site

FIG. 24B

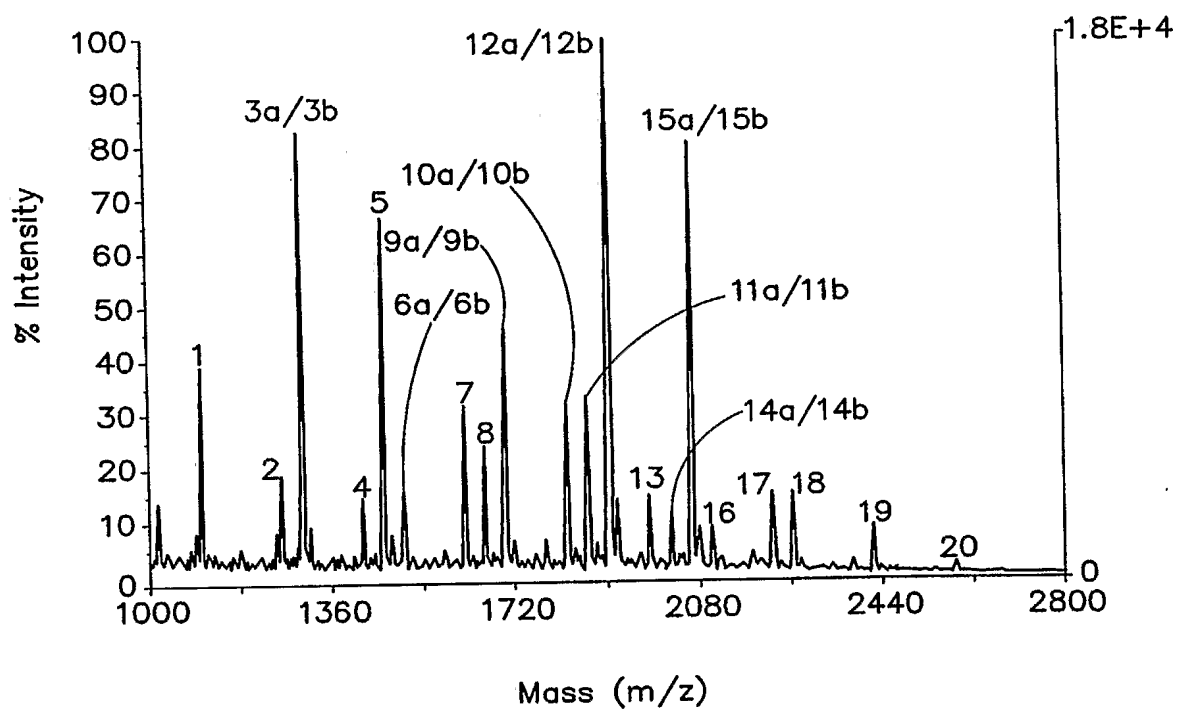
44/60

**FIG. 25A****FIG. 25B**

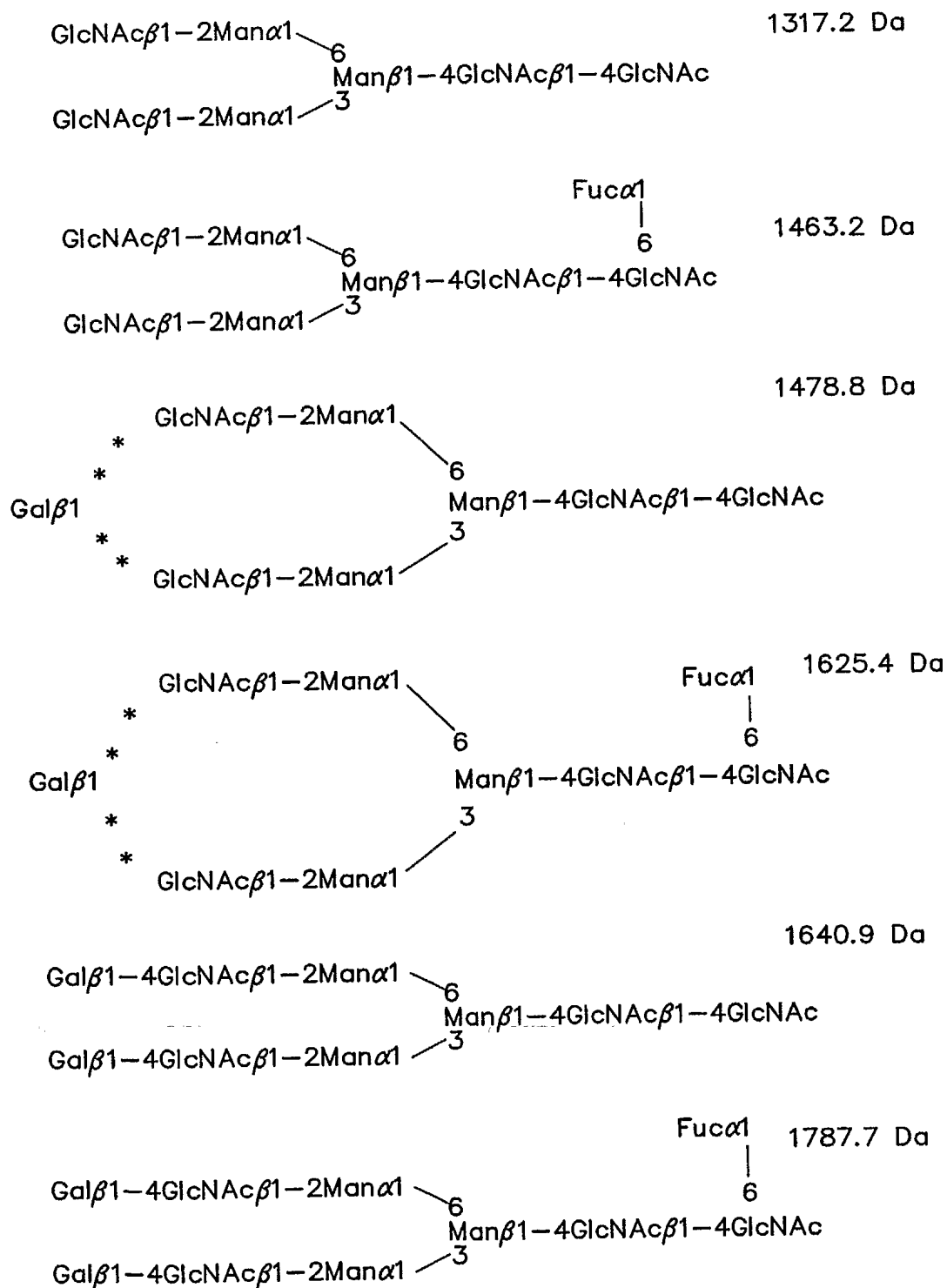
45/60

**FIG. 26**

46/60

**FIG. 27**

47/60

**FIG. 28**

48/60

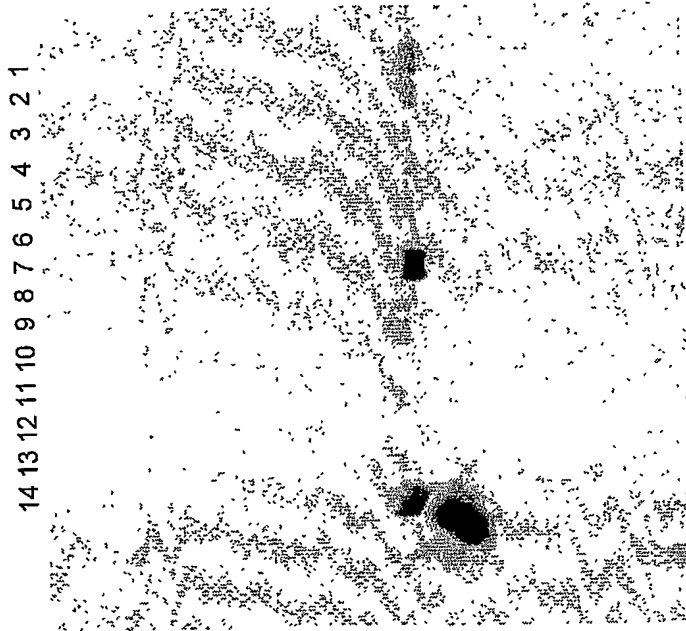


FIG. 29B

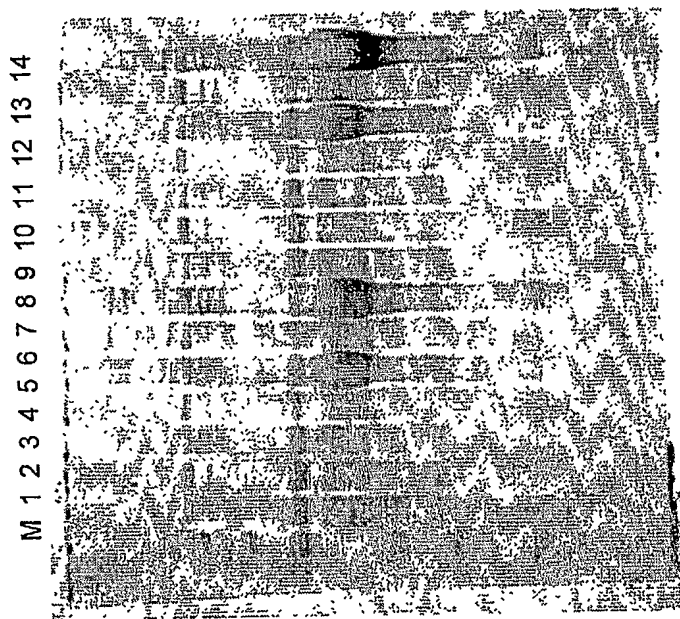
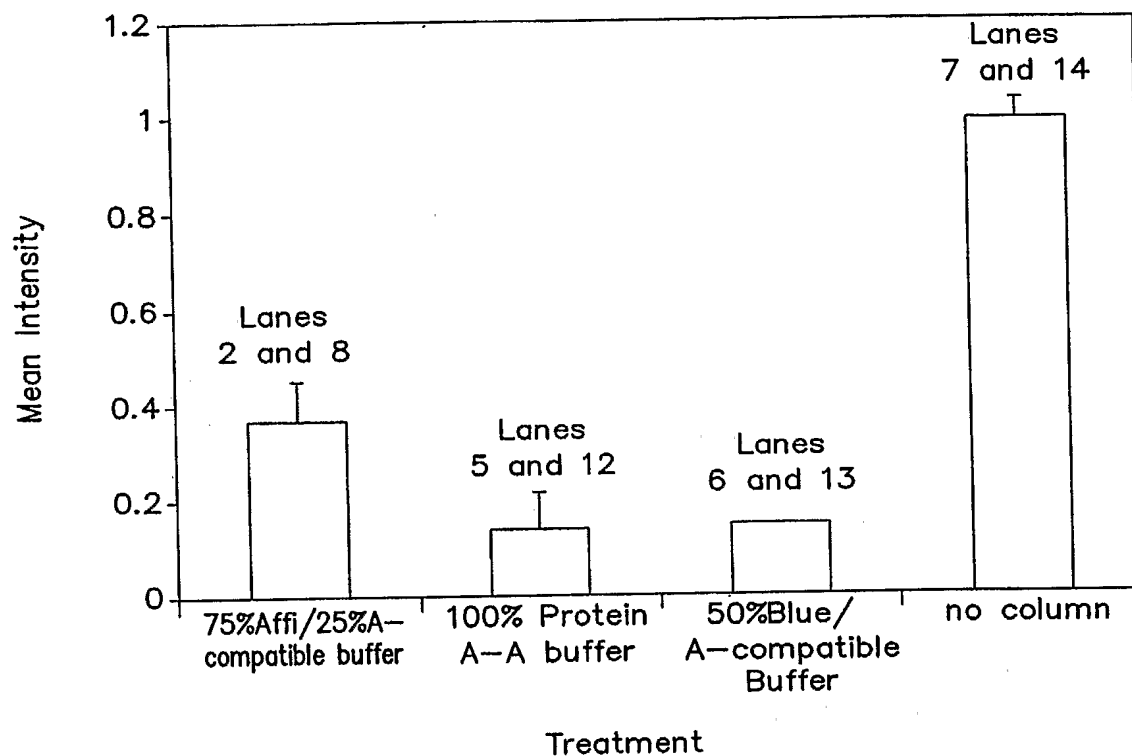


FIG. 29A

49/60

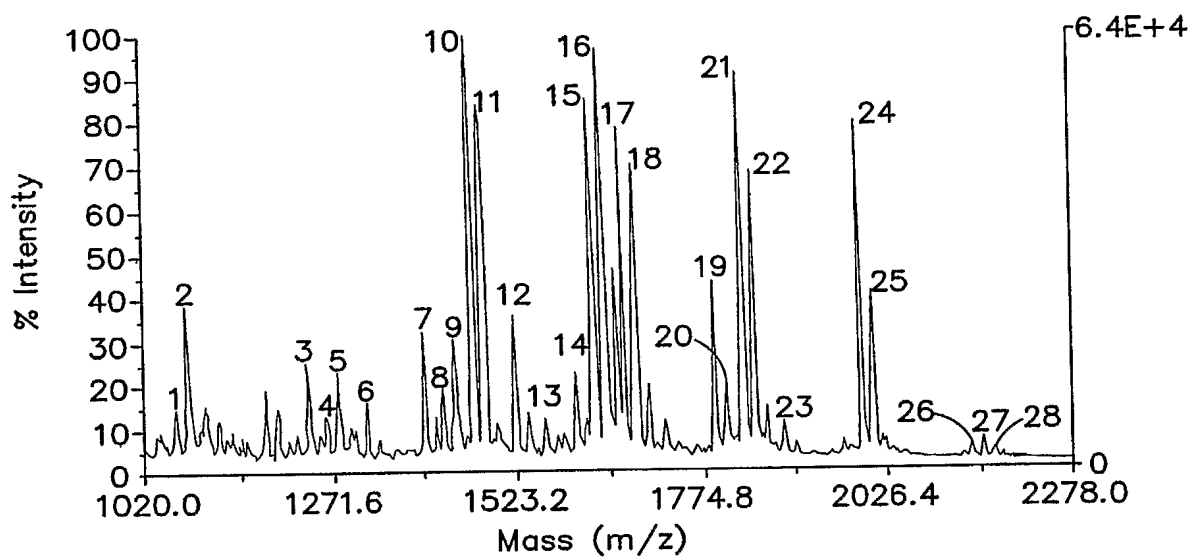
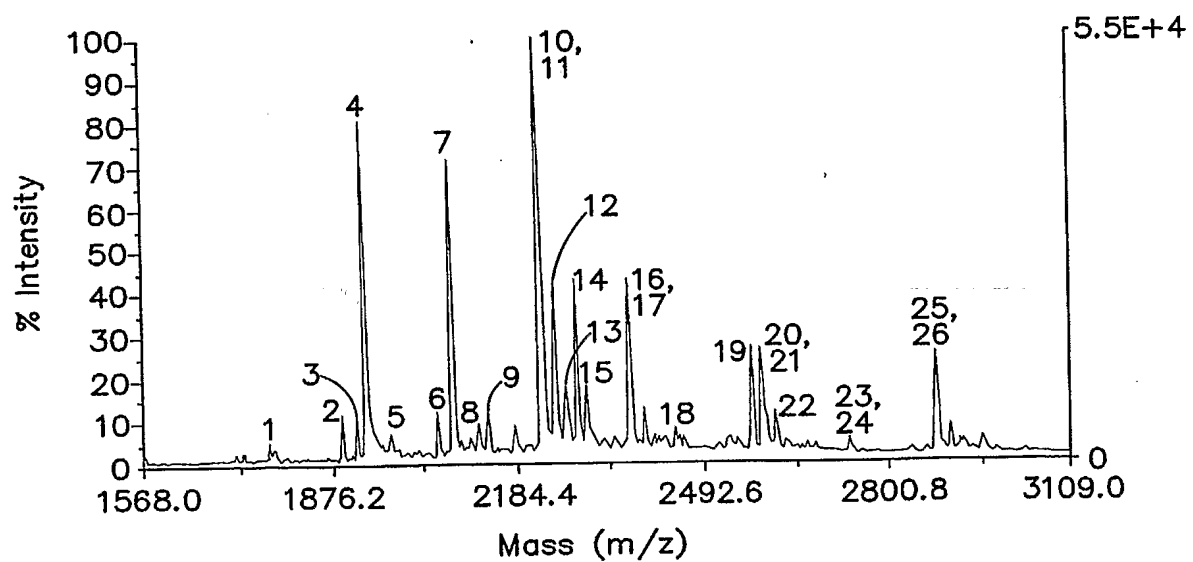
**FIG. 29C**

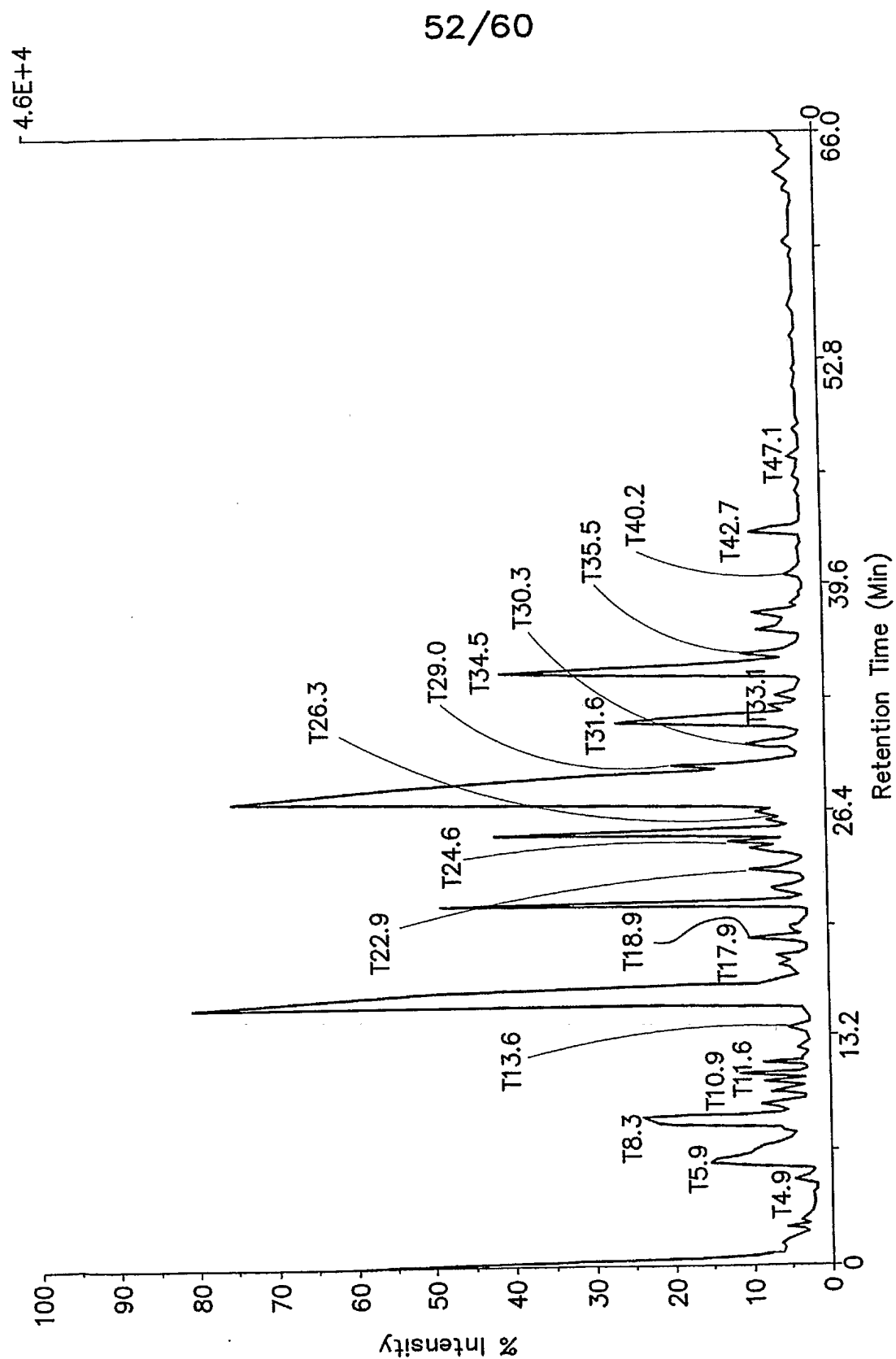
50/60

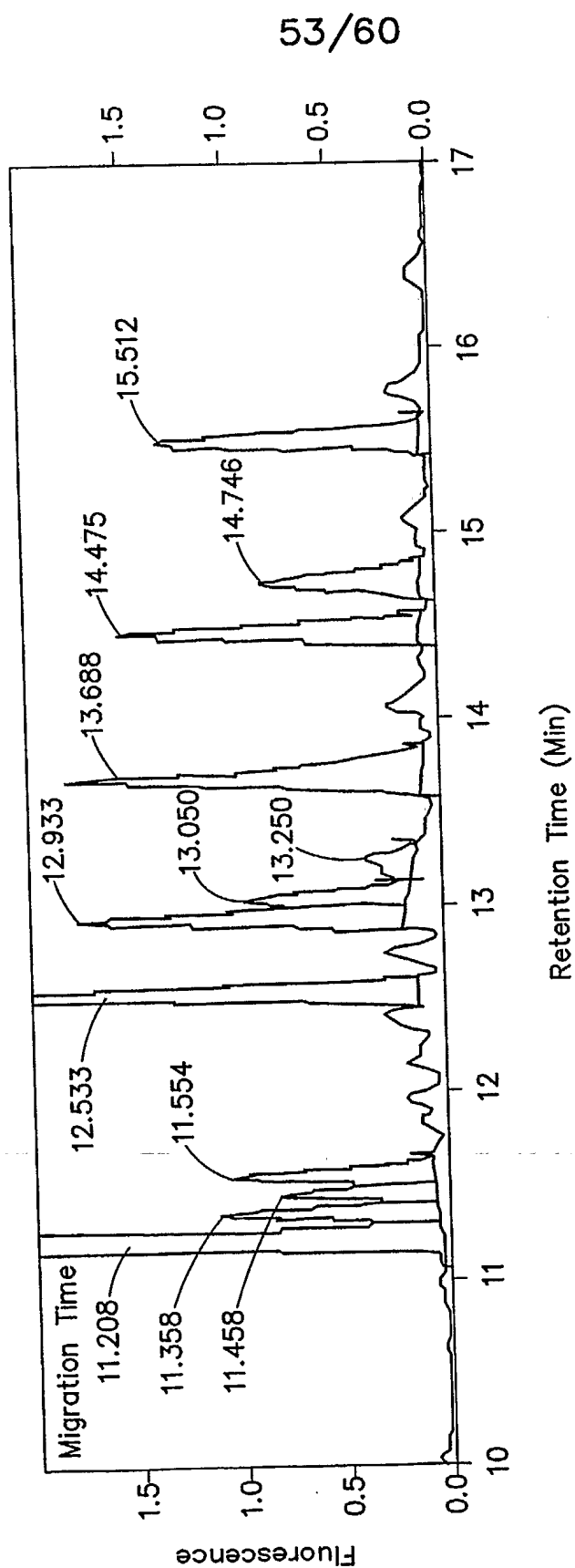


FIG. 30

51/60

**FIG. 31A****FIG. 31B**

**FIG. 32A**

**FIG. 32B**

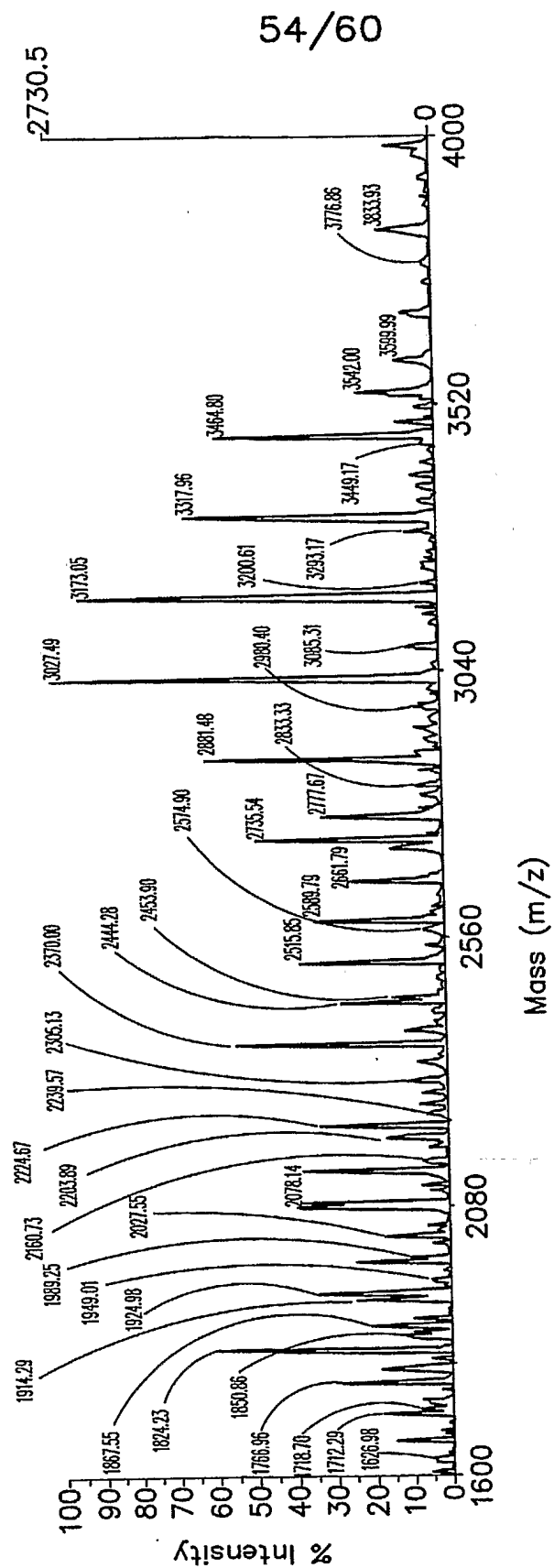
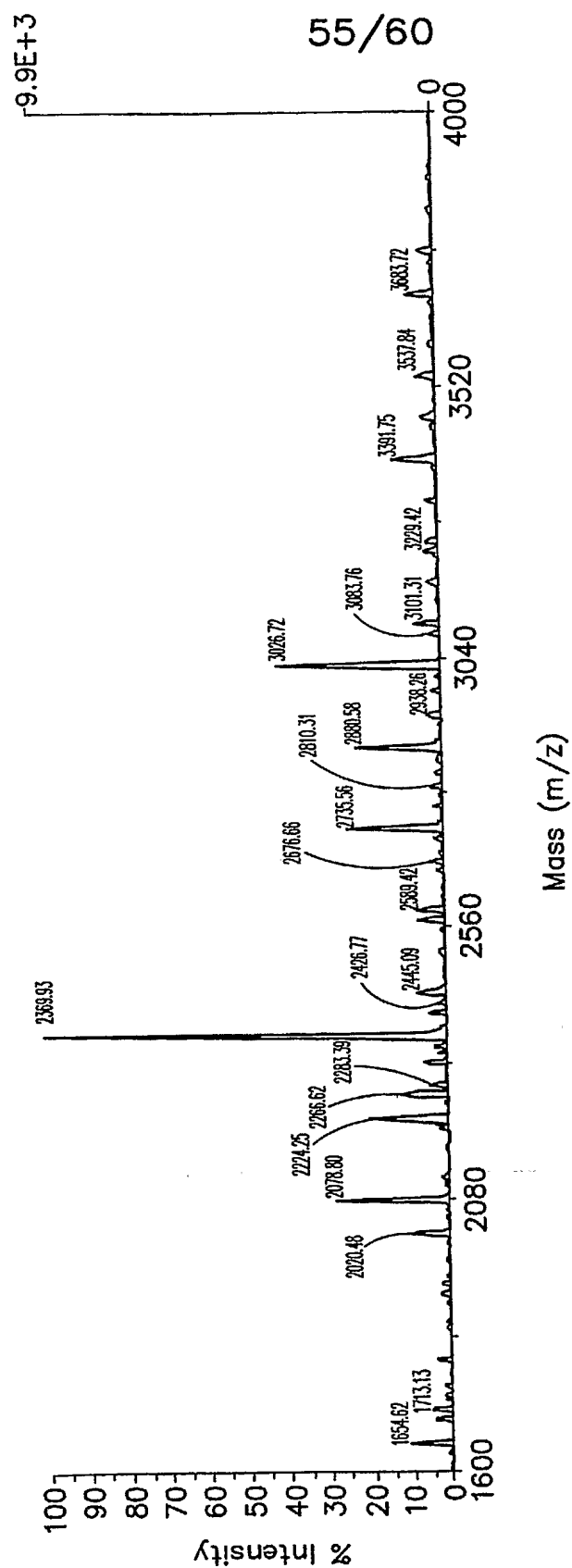
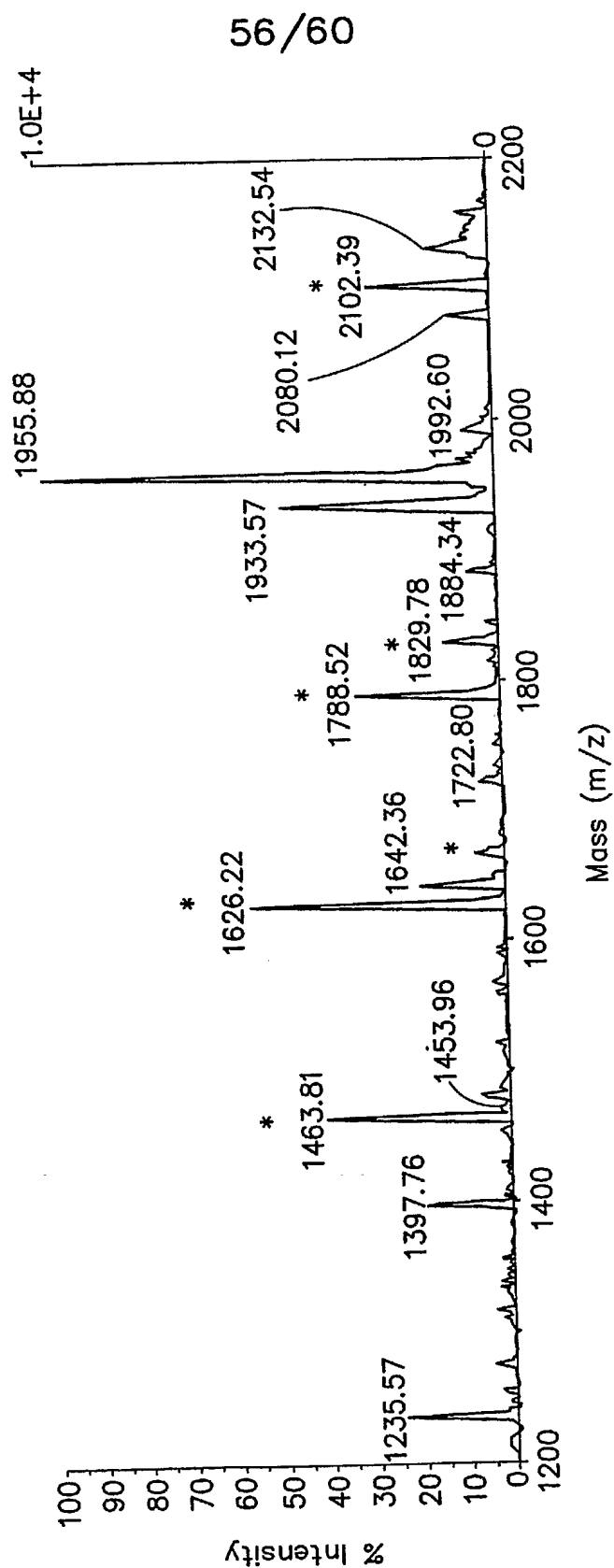
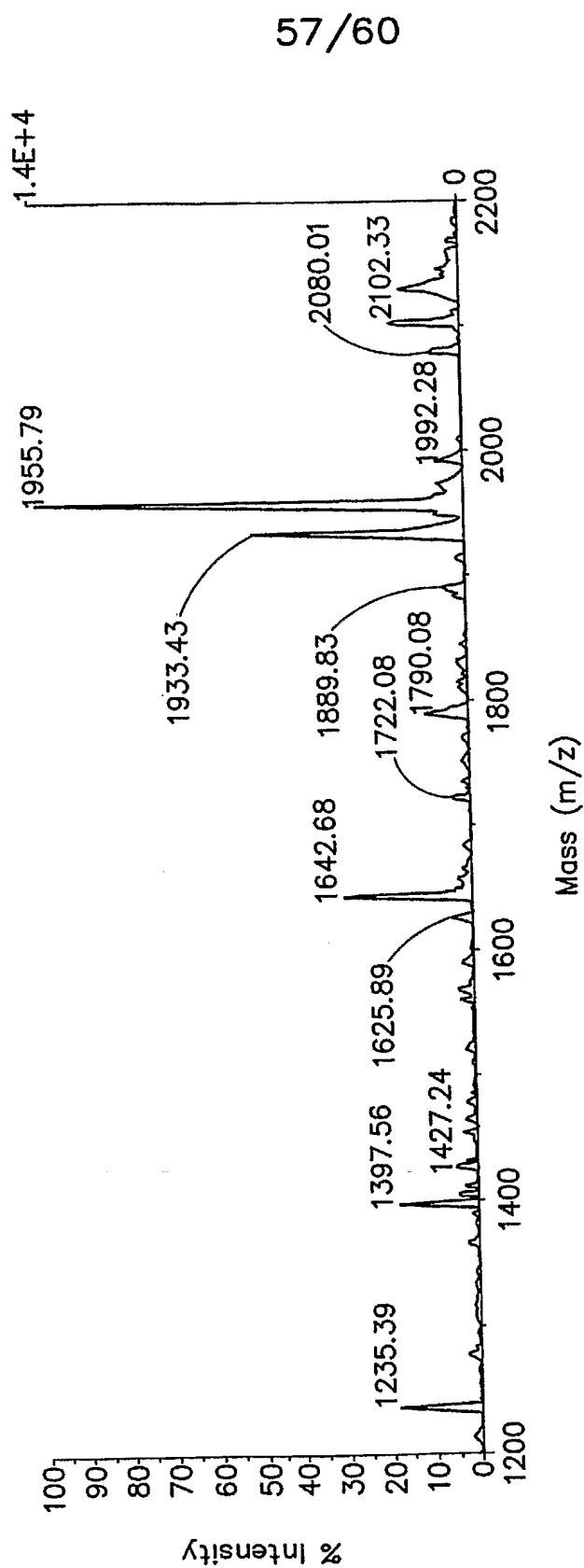


FIG. 33A

**FIG. 33B**

**FIG. 34A**

**FIG. 34B**

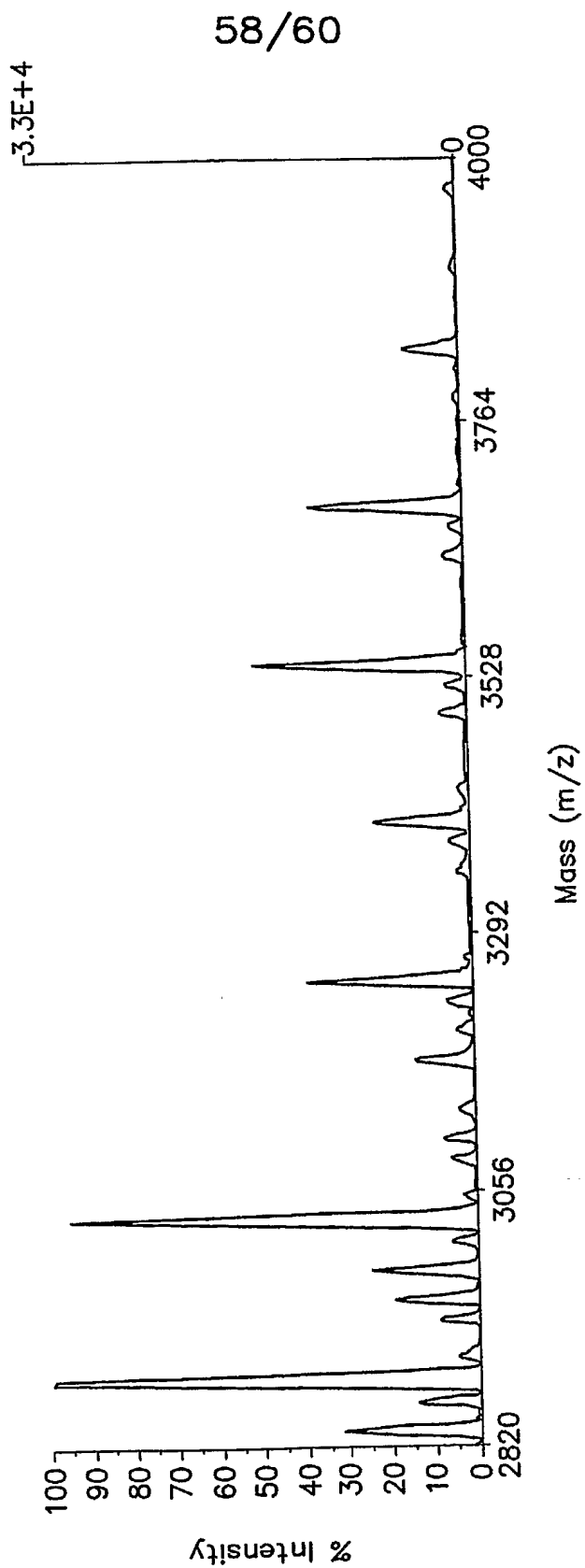


FIG. 35A

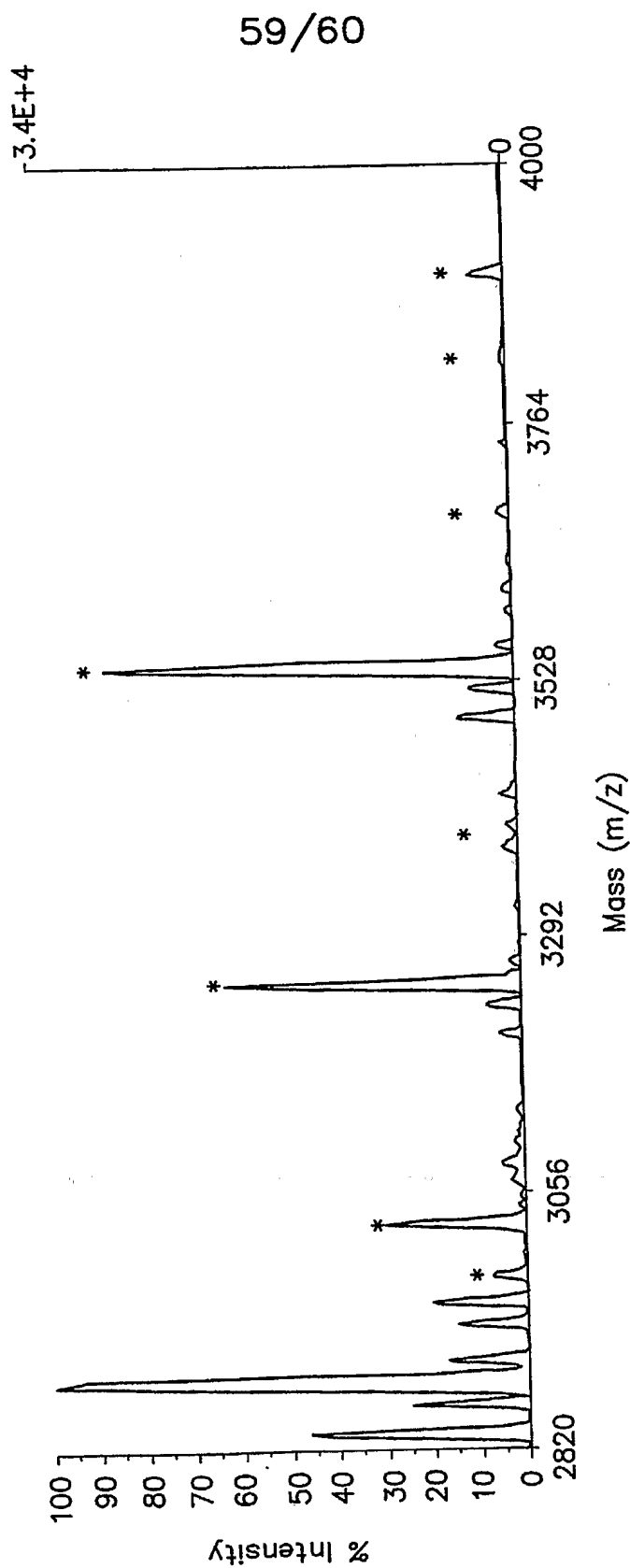
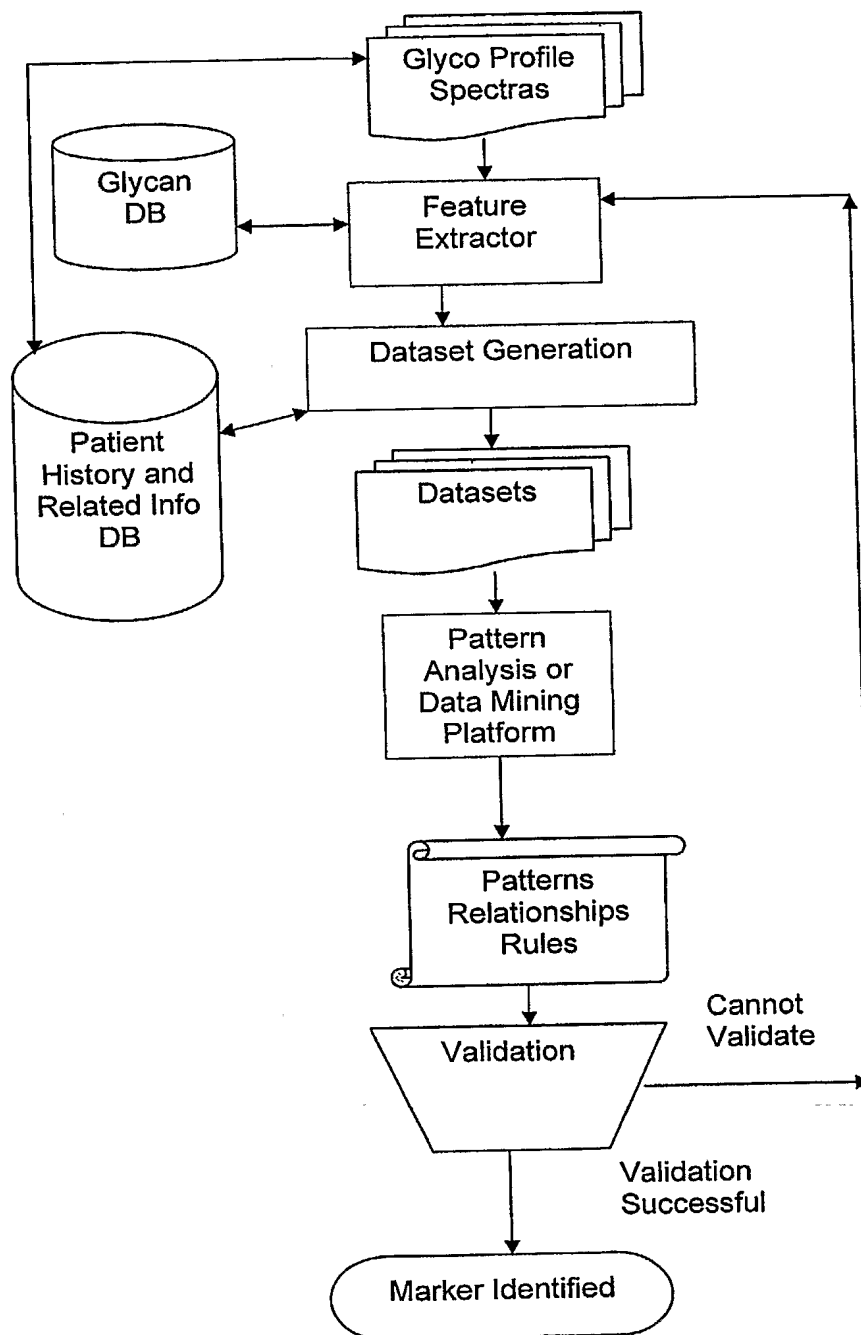


FIG. 35B

60/60

**FIG. 36**